

Self-assessment for informed study decisions in higher education

A design-based validation approach

Laurie Delnoij



Self-assessment for informed study decisions in higher education

A design-based validation approach

Laurie Delnoij

The research reported in this thesis was carried out at the Faculty of Educational Sciences of the Open University, Heerlen, the Netherlands



and under the auspices the research school Interuniversity Centre for Educational Research



© Laurie Delnoij, 2022

Printed by ProefschriftMaken | [Proefschriftmaken.nl](https://proefschriftmaken.nl)

Cover and lay-out: Vcreates by Vera van Laar

ISBN/EAN: 978-94-6423-682-8

All rights reserved.

Self-assessment for informed study decisions in higher education

A design-based validation approach

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van rector magnificus
prof. dr. Th. J. Bastiaens
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 1 april 2022 te Heerlen
om 16.00 uur precies

door

Laurie Emiel Chantal Delnoij

geboren op 8 oktober 1992 te Heerlen, Nederland

Promotor:

prof. dr. R. L. Martens, *Open Universiteit*

Copromotoren:

dr. J. P. W. Janssen, *Open Universiteit*

dr. K. J. H. Dirks, *Hogeschool Zuyd*

Leden beoordelingscommissie:

prof. dr. T. de Laet, *KU Leuven*

prof. dr. E. Braun, *Justus-Liebig Universität Gießen*

prof. dr. Th. J. Bastiaens, *Open Universiteit*

dr. K. Schlusmans, *Open Universiteit*

dr. R. Kappe, *InHolland Hogeschool*

Witse Laurie, went ich urgens aa begin, da sjei ich gewoon neet mieë oet.

~

Weetje Laurie, als ik ergens aan begin, dan stop ik er gewoon niet mee.

~

You know what, Laurie, when I start something, I just don't quit.

Merle Vandeberg (6)

April, 2021

Content

General introduction	3
Chapter 1 Predicting and resolving non-completion in higher (online) education – A literature review	16
Chapter 2 Predicting completion: The road to informed study decisions in higher online education	48
Chapter 3 Designing an online self-assessment for informed study decisions: The user perspective	78
Chapter 4 Towards an online self-assessment for informed study decisions – A mixed-methods validation study	92
Chapter 5 Do self-assessments for informed study decisions actually inform study decisions? A model for evaluating the consequential validity aspect	116
General discussion	143
Appendices	161
Summary	207
Samenvatting	215
References	223
Dankwoord	237
ICO Dissertation Series	243



General introduction



Non-completion in higher education

Non-completion is a problem for students, educational institutions, and for society at large, for numerous reasons (OECD, 2018, 2020; Vossensteyn et al., 2015). These reasons reach as far as student confidence and well-being, institutional reputation, and plain financial or time-related (return to investment) costs from all viewpoints: individual, institutional, and societal (Di Stasio & Solga, 2017; Oreopoulos & Petronijevic, 2013; Simpson, 2006; 2010). Completion is one of numerous terms mentioned as part of academic achievement or student success, and it is often referred to as persistence or retention. Its counterpart, non-completion, is also referred to as attrition or drop-out (Muljana & Luo, 2019; Rovai, 2003; Simpson, 2010; 2013; Vossensteyn et al., 2015). In this thesis we choose for the more neutral term 'completion', defined as *"meeting the requirements for certification of a course or program within a specified period of time"*.

In traditional higher education (i.e., face-to-face education in universities of applied sciences or research universities), non-completion is a large problem. Non-completion rates range from 17 to 47 percent within the first year following enrolment (i.e., based on figures of fourteen European countries, see Vossensteyn et al., 2015). In *online* higher education (i.e., blended and distance education), researchers even report non-completion rates between 50 and 98 percent (Aragon & Johnson, 2008; Levy, 2007; Morris et al., 2005; Patterson & McFadden, 2009; Simpson, 2013).

Despite considerable effort from institutions to prevent non-completion, the problem remains persistent, especially in *online* higher education (McGrath et al., 2014; Rovai, 2003; Vossensteyn et al., 2015). Prior to the Covid-19 pandemic, participation in online higher education already rose steadily during the last decade (Allen & Seaman, 2013; 2017; Allen et al., 2016; Seaman et al., 2018). During the pandemic, this increase grew even further. It is predicted that after the pandemic, forms of online and blended education will continue to play a more prominent and lasting role (Gomez Recio & Colella, 2020). Together, this results in a growing need to answer the following main question: *How can we clarify and effectively address non-completion in online higher education?*

Addressing non-completion

In order to enhance completion, it is important to have a clear understanding of the problem and the factors that are related to the issue first. Based on that, interventions can be developed targeting those factors that are likely to sort the most effect. Explaining non-completion has occupied researchers for years, especially in the context of traditional higher education. Several researchers studied single predictors of completion, or combined results of single predictors in review studies (e.g. Richardson et al., 2012; Robbins et al., 2004). Others studied integrated theoretical models, which combine predictors in order to explain completion. A well-known example is the Student Integration Model by Tinto (1975), which has been adapted over time (e.g. Neuville et al., 2007). From this body of prior work we can conclude that traditional cognitive factors such as prior education and scores on standardised ability tests are stable predictors of completion (Richardson et al., 2012; Robbins et al., 2004). Next to that, also non-cognitive

factors such as study behaviour and motivation appeared to be important factors for explaining non-completion (Allen et al., 2009; Richardson et al., 2012; Robbins et al., 2004). The first follow-up question arises here: *How do we influence these kinds of factors with interventions in order to increase completion rates?*

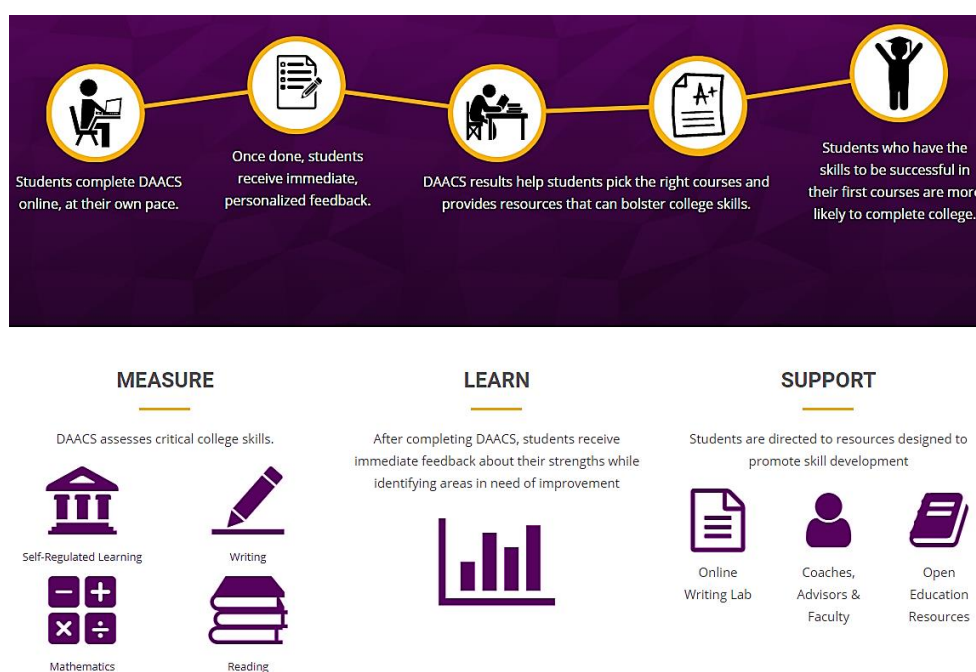
In a recent review about predictors and interventions for completion, Muljana and Luo (2019), emphasised the need to intervene early, even before student enrolment. Pre-enrolment interventions aim to improve the alignment between students' skills, motivation, and cognitive beliefs on the one hand and the 'demands' of higher (online) education on the other hand. They do so by raising prospective students' awareness and providing early remediation (Demulder et al., 2019; Fonteyne & Duyck, 2015; Nolden et al., 2019; Robinson et al., 1996). Various studies stressed the need for pre-enrolment intervention because students who end up not completing a course or program, often appeared to be lacking sufficient and/or timely information about whether their characteristics and abilities matched those required to succeed in an academic program (Germeijs & Verschueren, 2007; Hachey et al., 2013; Menon, 2004; Stinebrickner & Stinebrickner, 2014; Vossensteyn et al., 2015).

Online higher education is often operates according to a broader access compared to traditional higher education institutions, which makes pre-enrolment interventions even more relevant in that context. At the Dutch Open University (OUNL), for instance, the only admission requirement for bachelor programs is a minimum age of 18 years. No specific requirements are set regarding prior educational level, despite the university level. The openness and general flexibility of *online* higher education comes at a price, as exhibited by the higher non-completion rates. Taking into account this broader accessibility, pre-enrolment interventions should focus on increasing the number of students that meets the requirements for certification, without regulating admission or selection. This raises the second important follow-up question with regards to non-completion, specifically for pre-enrolment interventions: *How can we protect students from having unrealistic expectations and a frustrating study experience without setting (additional) entry requirements?*

Within these boundaries, pre-enrolment self-assessments seem a promising approach (Demulder et al., 2019; Lee et al., 2013; Muljana & Luo, 2019; Nolden et al., 2019; Pinxten et al., 2019). These assessments can provide adequate and personalised information, which is pivotal for prospective students to make a well-informed study decision, to stay motivated, and successfully complete their study (Kubinger, et al., 2012; Nicol, 2009; O'Regan et al., 2016; Pinxten et al., 2019; Tinto, 1999; Van Klaveren et al., 2019). Self-assessments prior to enrolment are informative advisory instruments, which induce self-examination (Hornke et al., 2013). In general, these instruments often involve diagnostic (proficiency, cognitive) tests and questionnaires on relevant non-cognitive variables. Often, they are concluded with feedback or (open-ended) advice to enable informed decision making (Soppe et al., 2019; Demulder et al., 2019; Nolden et al., 2019). Self-assessments are prolific, mainly in traditional higher educational practice.

For instance the Diagnostic Assessment and Achievement of College Skills (DAACS) (Bryer et al., n.d.), the Self Reflection Tool (Nolden et al., 2019), and Columbus (Demulder et al., 2019).

DAACS (see Figure 0.1.) is a diagnostic tool that measures prospective students' study readiness and provides them with immediate feedback about strengths and weaknesses along with links to resources on four areas: self-regulated learning, reading, mathematics, and writing. In the instruction of tests, measurements and their relevance are explained. Feedback aligned to the tests involves information on the obtained scores (visualized in three dots of which the (lack of) filling indicates the score degree) indicating prospective students' strengths and weaknesses. The focus of the feedback is on what prospective students can do in order to become better prepared and enhance their chances of success. To this end, the feedback entails video explanations with tips and links to online (open) resources and remedial courses.



Copyright 2021, DAACS c/o Jason Bryer, <https://daacs.net/>

Figure 0.1. *Diagnostic Assessment and Achievement of College Skills (DAACS)*

The Self-Reflection Tool (Nolden et al., 2019) is developed through a European collaboration. In this tool, prospective students can take tests on factors such as self-discipline, motivation, and learning strategies. Feedback based on these tests is aimed at raising the students awareness and self-reflection about their situation and study readiness (Nolden et al., 2019). The feedback provides information on obtained scores, in a traffic-light visualisation, with a general explanation of the test, and an advice on what the student could do in case they need help or for further preparation.

Columbus, a similar instrument implemented in Flanders (Demulder et al., 2019), includes tests on three categories: *Who am I* (e.g. motivation and study strategies), *What do I know* (i.e. cognitive skills like numerical skills and reasoning skills), and *What do I want* (e.g. interests). Feedback to prospective students in Columbus involves information on the obtained subtest scores compared to successful students in the first year after enrolment, and an advice for further preparation (Broos et al., 2018; 2019). Such feedback is presented to support prospective students in making well-informed study decisions (Nolden et al., 2019; Van Klaveren et al., 2019) and possibly leads to early remediation (Broos et al., 2018; 2019; Muljana & Luo, 2019). In turn, the aim is that this leads to a decent start and enhanced subsequent study success in higher (online) education (Kubinger, et al., 2012; Nolden et al., 2019; O'Regan et al., 2016; Van Klaveren et al., 2019).

In the Netherlands (the context of the present research), self-assessments are often part of a broader orientation activity called '*study choice check*' (in Dutch: *studiekeuzecheck*). Traditional higher educational institutions are obliged by Dutch law (Quality in Diversity Law, 2013) to offer such orientation activities, aimed at providing prospective students insight into the study level, content, and environment. The actual organization of the study choice check activities is up to each institution (Soppe et al., 2019). Self-assessments in this context entail tests on, for instance, motivation, ability beliefs, interests, and time management (Soppe et al., 2019). In most cases, the self-assessment is combined with other activities such as an online teaser course or class, a day on campus or a meeting with lecturers, and concluded with an advice regarding a student's fit with the programme (Knuiman & Kappe, 2017; Soppe et al., 2019).

The impact of these pre-enrolment activities is potentially far-reaching, for both the individual (student) in terms of decision-making and progress, and for the institute in terms of enrolment and success rates. For instance, an evaluation of the *study choice check* (in which self-assessment is one component) at a (traditional) Dutch University of Applied Sciences showed that students receiving a negative study advice more often decided not to enrol, in comparison to those with a sufficient or positive advice (Kappe & Knuiman, 2019). After implementation of the study choice check, non-completion figures in that context generally decreased (Kappe & Knuiman, 2019). It must be noted, however, that in this evaluation differences between programmes were found and results could not be compared to a control group of students who did not take part in the study choice check. Another study, by Van Klaveren et al. (2019), showed that providing students with feedback on expected success rates increased enrolment with about 25%, even though it did not reduce first year dropout.

Considering that self-assessments prior to student enrolment seem a promising approach but are mainly studied in traditional higher education (Fonteyne & Duyck, 2015; Kubinger et al., 2012; Muljana & Luo, 2019; Nolden et al., 2019), the focus in the present thesis is to develop a similar instrument for prospective students in higher *online* education. In light of the accessibility of *online* higher education, feedback (or advice) resulting from self-assessment and its impact should be justified and fair. After all, we do not want to discourage students unnecessarily. Therefore, it is important that impact of self-assessments in the context of study decisions is theory- and data-driven (Demulder

et al., 2019; Nolden et al., 2019). Currently, pre-enrolment (self-)assessments are rarely transparently designed or validated (Niessen & Meijer, 2017). As access to higher education requires the best possible decision making support, we argue that these assessments should be validated as fully and explicitly as (summative, high-stakes) standardized assessments (Wools et al., 2010).

The quest for design-based validation

Solving complex problems such as the non-completion problem in education, requires design based research (DBR) in the context for which a solution is demanded and in close and systematic collaboration with various stakeholders (e.g. students, practitioners, policymakers) (Van den Akker et al., 2013; Collins et al., 2004; Martens, 2018; Muljana & Luo, 2019). Design-based research is interventionist (involves some sort of design), takes place in naturalistic contexts, and is iterative (Barab & Squire, 2004; Bell, 2004). Stages of analysis, design, and evaluation are iterated until an appropriate balance between what is intended and what is realised has been achieved (Van den Akker et al., 2013).

Accordingly, this thesis describes studies in the stages of analysis, design, and evaluation, with overarching ongoing reflection on and revision of a prototypical self-assessment aimed at informed study decisions (Barab & Squire, 2004). The self-assessment at stake in the present thesis is non-committal, non-selective, but diagnostic: the aim is to enable informed decision-making (food for thought), and to encourage prospective students to start well-prepared (feedback for action). These aims pose high demands on assessment validity, i.e. *do the test scores, the feedback provided alongside, and prospective students' interpretations thereof, all match the proposed use of the assessment?*

In order to meet these demands, the design-based development process of the self-assessment (SA) involves evaluation of five sources of validity evidence, in line with modern validity theories (AERA et al., 2014; Messick, 1989). The five validity evidence sources relate to corresponding validity aspects: *content, predictive, internal structure, process* and *consequential* validity.

Based on the Standards for Educational and Psychological Testing (AERA et al., 2014; Beckman et al., 2005; Cook et al., 2014) these validity aspects are defined as:

- Content aspect: the extent to which the test content accurately represents the content domain;
- Process aspect (response processes): the fit between what tests or test items intend to measure and the experience appear to have and considerations they appear to take into account when responding;
- Internal structure aspect: the degree to which test items reflect coherent dimensionality, both on theoretical and statistical ground;
- Predictive aspect (relations to other variables): the relative performance of test scores in predicting (supposedly) related variables;

- Consequential aspect: interpretations of and actions following test result and the extent to which these are in line with intended uses of a test.

Figure 0.2. illustrates the various sources of validity evidence and their relevance at the various stages of the design-based development of the SA. The **analysis stage** focuses on answering the following questions: What factors are related to completion and, therefore, should be the target of the self-assessment, and how can these factors be measured in a self-assessment in order to detect students at risk of non-completion? This stage thereby focuses on collecting evidence for *content*, *internal structure*, and *predictive* aspects of validity. These aspects are evaluated to establish a model of predictors of completion, for the specific context in which the self-assessment is developed. Tests on predictors of completion are to be included as subtests for the self-assessment.

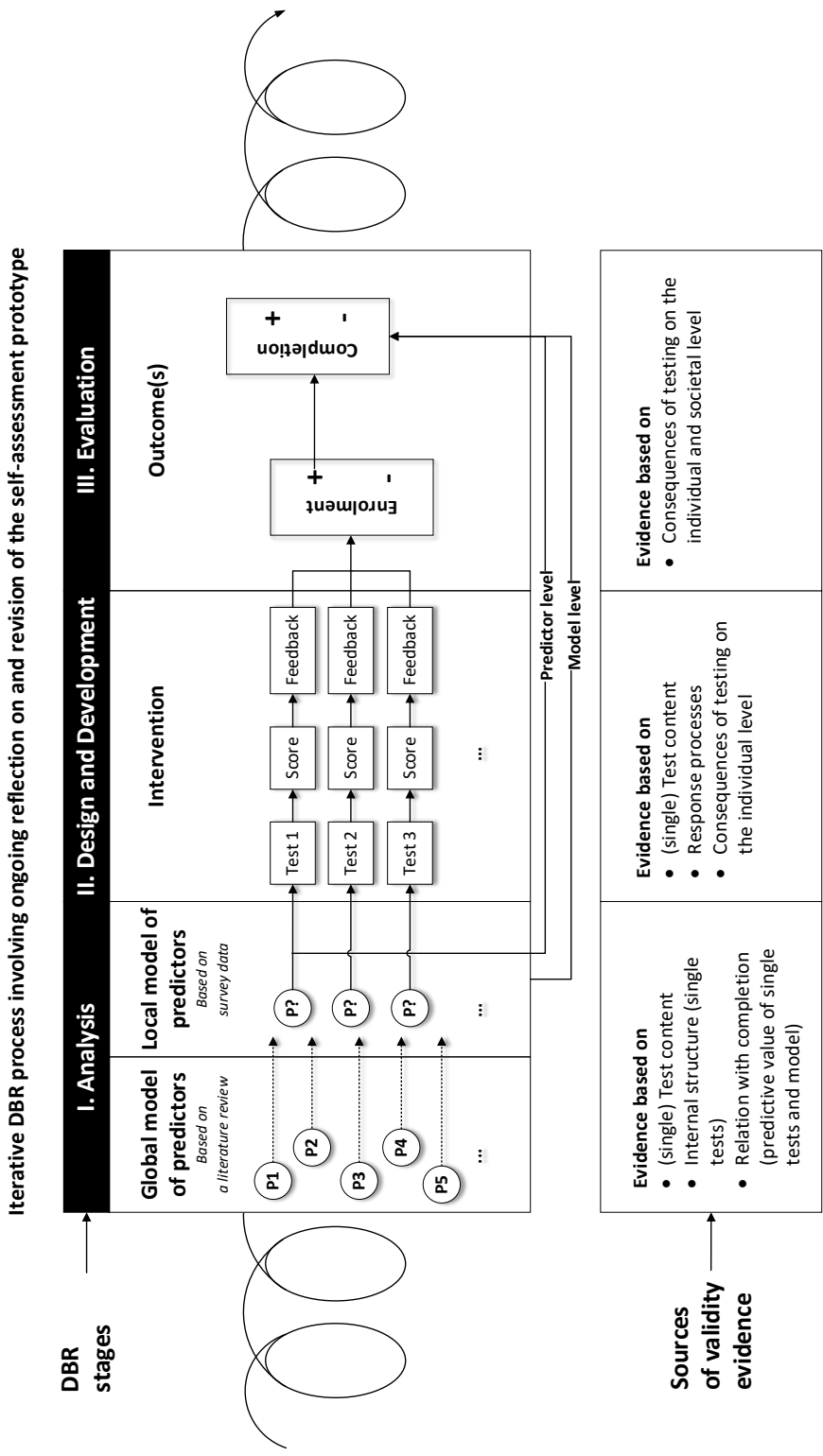


Figure 0.2. Design-based validation process of a self-assessment for informed study decisions

In the **design stage**, the focus will be on context- and target group specific requirements for the self-assessment. For instance, what should the feedback based on the assessment scores look like in order for prospective students to make sense of it?

In this stage, evidence on the *content* aspect of validity is to be supplemented from a user perspective. From both a DBR as well as a validity perspective, involving stakeholders (i.e. experts and/or those undergoing or working with the assessment procedure) in the design process is of importance (Barab & Squire, 2004; Beckman et al., 2005). In the regard it is important to note that content is not limited to the subtests included in the self-assessment, but also involves the feedback aligned to those subtests. To determine what the feedback aligned to SA subtests should look like, potential users of the instrument should be consulted as well. After all, research has shown that if prospective students do not perceive the feedback to be useful, it becomes less likely that they will take into account the information in making a study decision (Mittendorff, 2015; Warps et al., 2017).

Also in this stage, evidence on the *process* aspect of validity is to be evaluated. This aspect focuses on users' test taking strategies, actions based on, and thought processes regarding (a) test (items) (Beckman et al., 2005). Little is known about how prospective students proceed through self-assessments for study decision-making, as validity research (in general) tends to mainly focus on content, internal structure and predictive aspects (Cook et al., 2014; Kreiter, 2016). In regard to response processes, a general point of concern is that self-assessments, i.e. self-report measures, may be subject to various kinds of measurement errors, due to inaccurate self-perceptions (Dunning et al., 2004) or socially desired answers (Niessen et al., 2017; Viswesvaran & Ones, 1999). Therefore, users' test-taking strategies and reactions on the self-assessment need to be examined to determine whether the self-assessment is used as intended.

Furthermore, in the design stage, the *consequential* aspect of validity will be evaluated. This aspect pertains to anticipated and unanticipated consequences – both positive and negative – of measurement on an individual and societal level (Cook et al., 2014; Downing, 2003; Goodwin & Leech, 2003; St-Onge et al., 2017), which can support or challenge the soundness of score interpretations and actions based upon them (Beckman et al., 2005). Cook et al. (2014) especially argue that greater emphasis is required on describing and defending the decisions and actions following score interpretation, i.e. the consequential aspect of validity. In the context of study decision support tools, it appears such evaluation is often implicit or lacking (Niessen & Meijer, 2017). With regard to the SA, anticipated consequences range from individuals' interpretations of the scores and feedback to the decision on whether or not to enrol and the success after enrolment. Investigating the impact of the SA on a larger (societal) scale requires it to be fully available for prospective students (i.e. mainstream deployment). Before doing so, investigating consequences on an individual level helps to shed light on the question whether anticipated effects (e.g. intention for further preparation) are evoked as intended. This might indicate some final, yet critical changes before full implementation of the SA.

Although some kind of (prototypical) evaluation already takes place in the design stage, in the **evaluation stage**, the purpose of evaluation is rather summative (Kane,

1992). After implementation of the SA, the focus is on evaluating the *consequential* aspect of validity, not only on the individual level, but also on the societal level. On an individual level, the self-assessment might affect study choice certainty (Soppe et al., 2019) and/or lead to postponing an enrolment decision and taking remedial action (Broos et al., 2018; 2019). On a societal level, there might be an impact of the SA on enrolment and completion rates (Kappe & Knuiman, 2019; Van Klaveren et al., 2019). In addition, determining the effectiveness of interventions such as the self-assessment, does not only involve an evaluation of its impact on those outcome measures, but should also take into account other factors as assessment fairness (Kreiter, 2016; Xi, 2010), cost-effectiveness, and scalability (Kraft, 2020).

Investigating these sources is not a ‘once and for all’ activity, but one that requires continued attention, as student populations and/or educational practice may evolve over time (Messick, 1989; Royal, 2017). As indicated by the ongoing cycles in Figure 0.2., the development process involves ongoing reflection on and revision of the self-assessment prototype.

The current thesis

Objective and contribution

The main objective of the research presented in this thesis is to evaluate five sources of validity evidence for the purpose of designing a self-assessment for informed study decisions in online higher education. With this objective, we address the lack of empirical evaluation of such self-assessments aimed at informed decision-making. Although DBR is practice-oriented and inherently context-specific, it also aims to contribute to scientific theory building, on three different levels (Edelson, 2002):

1. Domain theories are descriptive and tell us something about a generalization of some kind of problem analysis. The present research contributes to domain theories about completion in online higher education (e.g. by determining predictors of completion) and theories about the study decision process (e.g. by gaining insight into how prospective students proceed through this process).
2. Design frameworks are prescriptive and indicate the requirements or characteristics of a particular design for particular purposes. The present research will result in an indication of what a self-assessment for informed study decisions should entail or look like, for it to actually *inform* prospective students’ decisions and support them in preparing for studying in online higher education.
3. Design methodologies are also prescriptive, though not focused on the design itself, but on the design procedure. The present research adds to the literature by providing a hands-on example of applied validation studies, which – so far – tend to focus solely on high-stakes assessments (i.e. selection, pass/fail or grade), standardized tests, and predominantly in the context of health professions (Cook et al., 2014; Wools et al., 2010).

Outline

The thesis involves three parts, following the design-based research stages. The first part examines predictors of non-completion in higher (online) education and interventions to enhance completion in that context. This part aims to analyse and explain the non-completion problem in the general context of higher education. The second part focuses on the design and development of the self-assessment for informed study decisions for the specific target group and within the specific context (OUNL). In this part (design stage), an evaluation will take place, focused on refining the prototypical self-assessment. After mainstream implementation (evaluation stage), evaluation becomes summative, focusing on the impact of the self-assessment in practice.

Analysis stage

Chapter 1 describes a *systematic literature review* focused on predicting and resolving non-completion in higher education. The results entail two overviews. First, an overview of predictors, in which their predictive consistency (stability of results across included studies) and modifiability (i.e. the extent to which predictors can be influenced by interventions) is taken into account. The second overview entails characteristics of (effective) interventions aimed at increasing completion in higher (online) education. This overview provides insight into the extent to which these interventions focus on the most consistent and modifiable predictors, as established in the first part of the literature review. In other words, through this literature review, we establish a global model of predictors of completion. Thereby, the study provides the first sources of validity-evidence, for the content and predictive value of the self-assessment.

Chapter 2 reports on an *empirical study* in which we examine evidence on content, predictive and internal structure aspects of validity for consistent modifiable predictors of non-completion, selected from the review in Chapter 1, in the specific context for which the self-assessment is designed (i.e. the OUNL). Furthermore, we investigate the resulting predictors on their combined classification accuracy (i.e. to what extent do the predictors together accurately distinguish completers from non-completers?). The result of this study is a local model of predictors of completion and the first set of prototypical subtests of the self-assessment.

Design and development stage

Chapter 3 focuses on the perspective of potential users of the self-assessment – prospective students of higher *online* education. The fact that literature and predictive analyses suggest certain variables as relevant to be tested in the self-assessment, does not mean that potential users see that relevance as well. If they do not perceive the tests to be relevant, the chance that they will deliberately use the self-assessment and the information they can gain from it for their study decision becomes less likely. In addition, the content of the self-assessment is not limited to its subtests, but also involves the content of the feedback provided aligned to the obtained scores on these subtests.

In a *user study*, prospective students are asked what tests they would expect in a self-assessment during their orientation for studying in higher education and what feedback information they would expect aligned to their obtained scores. Based on this study, additions to the prototypical set of subtests in the self-assessments are proposed and the content of the feedback is further established.

Chapter 4 involves the first step in the development and validation process in which prospective students actually take the self-assessment. After establishing satisfactory results regarding content, internal structure and predictive aspects of validity in the previous studies and additional analyses, the focus is shifted towards the process and (individual) consequential aspects of validity in this study. In a *qualitative in-depth study*, prospective students take the self-assessment in an observed think-aloud mode. Before and after taking the self-assessment, they are interviewed on their expectancies of and experiences with the self-assessment. Resulting from this study are insights for the process aspect of validity as in prospective students' test-taking strategies and reactions on the subtests of the self-assessment. Additionally, this study provides insight into the individual consequences of testing as in, the impact of the self-assessment on prospective students' study choice certainty and intentions for further orientation and preparation. Based on this study and by final refinements, the self-assessment is assembled for '*mainstream deployment*'.

Evaluation stage

In **Chapter 5**, we present the results from the self-assessment after going in 'full release'. In an *explanatory evaluation study*, prospective students take the self-assessment in an authentic situation of orienting towards a course or study program at the OUNL. After taking the self-assessment, they are asked about the impact of the self-assessments on their study choice certainty. In addition, their obtained scores and answers to the evaluation survey are linked to enrolment behaviour. Based on this evaluation, recommendations for further development of the SA and its implementation in educational practice are highlighted.

The thesis concludes with a **General Discussion** which provides an overview of the main findings, recommendations for self-assessments aimed at informed decision making and future research.



Chapter 1

▮ Predicting and resolving non-completion in higher (online) education – A literature review

This chapter is based on: Delnoij, L. E. C., Dirkx, K. J. H., Janssen, J. P. W., & Martens, R. L. (2020). Predicting and resolving non-completion in higher (online) education – A literature review. *Educational Research Review*, 29, 100313. <https://doi.org/10.1016/j.edurev.2020.100313>

Non-completion in higher education is a persistent problem and even worse of a problem in higher online education. Although there is a lot of research on predictors of non-completion, less is known about what interventions resolve the non-completion problem and to what extent these interventions focus on relevant predictors of non-completion. To close that gap, the literature was systematically reviewed with a twofold aim: 1. Identify modifiable predictors of non-completion in higher (online) education 2. Investigate characteristics of effective interventions to reduce non-completion in higher (online) education. Results showed that study- or learning strategies, academic self-efficacy, (academic) goals and intentions, institutional or college adjustment, employment, supportive network, and faculty-student interaction are modifiable consistent predictors of non-completion. Coaching, remedial teaching, and peer mentoring are promising interventions to resolve the problem of non-completion in higher education. Interventions aimed at increasing completion rates are limited in targeting relevant modifiable predictors of non-completion.

1.1. Introduction

Non-completion is a problem for students, educational institutions and society at large for various reasons that go beyond the straightforward issues of efficiency and effectiveness, such as effects on students' confidence and institutional reputation (Simpson, 2006, 2010; Vossensteyn et al., 2015).

Completion in the current research is defined as: *meeting the requirements for certification related to a course or program*. Completion rates thus indicate the proportion of students enrolling in a course or program and meeting the requirements for certification, within a specified period of time. For this literature review, we look at completion rates within the first year of higher education, as most students who do not complete a course or program tend to dropout during or immediately after the first year (Simpson, 2010; Tinto, 2012; Willcoxson et al., 2011).

Despite the fact that the non-completion problem is on the agenda of numerous universities and the considerable effort from institutions to prevent non-completion, the non-completion numbers are still eminent (Vossensteyn et al., 2015). In the context of traditional higher education, non-completion rates range from 17% to 47% (i.e. based on figures of 14 European countries, see Vossensteyn et al., 2015). Non-completion in the higher online educational context (e.g. blended and higher distance education) appear to range from 78% to around 99% (Simpson, 2013). However, non-completion figures are quite diverse, as they are highly dependent on enrolment policy and definitions of completion, and different methods are used to calculate these numbers (Rovai, 2003; Simpson, 2010, 2013; Vossensteyn et al., 2015). On the whole, non-completion is worse of a problem in the higher online educational context (e.g. blended and higher distance education). First, because the numbers of non-completion are greater, but also because online education has grown tremendously over the past decade (Seaman et al., 2018). The higher online educational context differs from the traditional higher educational context in various respects. Higher online education is delivered fully online or in blended formats (i.e. a combination of online and face-to-face). This generally means more flexibility in the sense that studying becomes largely place, time, and pace independent (Wedemeyer, 2010). As a result, the higher online educational context generally attracts students who combine a study with other activities (e.g. a job, family or community obligations). This means that higher online education generally, though not exclusively, involves adult learners. It is important to take into account that the ambitions of students in higher online education may not be degree-oriented. In this respect, it is important to distinguish between the concepts of completion and study success. Though there is little evidence on this issue, there is research suggesting that not all students in higher online education start a course or program with the intention to obtain a certificate (Henderikx et al., 2017; Schlusmans & Winkels, 2017). Schlusmans and Winkels (2017) for instance, have reported that in a distance university context, approximately one-third of the students do not aim to obtain a diploma. It might be that these students, enrolling in a course or program without completion still have attained particular learning goals. Therefore, they cannot be said to have failed or been unsuccessful. For this reason, we here use the more neutral terms completion and non-completion in higher (online)

education, rather than a term like ‘study success’. However, even taking this into account, completion rates in higher online education demand improvement (Rovai, 2003; Schlusmans & Winkels, 2017). Though to a certain extent, non-completion is inherent in higher (online) education, current figures are still seen as problematic, as evidenced by the many studies and initiatives in higher (online) education to explain and/or reduce non-completion. One of the reasons that non-completion rates are still poor might be that initiatives taken to reduce non-completion do not focus on relevant variables explaining or predicting non-completion and this will be the focus of the current review.

There are two determinants in the completion rate equation: the number of students meeting the requirements (numerator) and the number of students enrolling (denominator). In theory, then the odds of completing (completion rates) will improve when either more students meet the requirements under equal enrolment numbers, or the number of students meeting the requirements remains the same under reduced numbers of enrolment. The latter effect might stem, for instance from, a communication and admission policy that increases the chances that those enrolling will meet the requirements. Increasing the number of students meeting the requirements might be achieved by increasing the effectiveness of the learning process, for instance, by more adequate instruction, tutoring, and guidance. In other words, interventions to increase completion rates are possible both prior to and after enrolment. Interventions prior to enrolment might be, for instance, a trial studying procedure for prospective students, or diagnostic assessments. After enrolment, there is a wide variety of possible interventions, for example, a counselling trajectory with a student advisor, training in effective learning strategies or curriculum changes to enhance completion rates. In line with this completion rate equation, Elffers (2018) refers to a trilemma involving accessibility of education, quality of education, and study success. According to this trilemma, study success can be increased by reducing the accessibility of education on one hand or increasing the quality of education on the other hand. It goes without saying that accessibility constitutes a sensitive ethical issue, which especially in the context of open education, is subject to certain constraints.

Before effective and efficient interventions can be designed and researched, it is important to have a comprehensive picture of the modifiable factors that predict non-completion, so that interventions can be developed targeting those factors that are likely to sort most effect. To inform the future development of interventions aimed at further improvement of completion rates, a literature review was conducted. The following two research questions guide this endeavour¹ :

1. Which (modifiable) variables are most strongly related to non-completion in higher (online) education? As several review studies have already tried to summarize the vast amount of studies aiming to explain the non-completion

¹ In the next sections in this Chapter we refer to higher education as the context of this research, by which we thus mean traditional higher education, but also higher online education.

problem, this literature review addressing this question will build on these review studies.

2. What are the key characteristics of interventions that proved effective in increasing completion rates, in which context and to what extent? To our knowledge, no systematic review of intervention studies has been done yet. It will be interesting to relate the answers to both questions, to see to what extent interventions developed so far, actually target the variables that the review studies indicate to be most strongly related to non-completion.

In the next section a detailed description of the literature search, selection and data synthesis will be provided.

1.2. Methods

1.2.1. Search and selection

To find relevant articles in line with the aim of this review we consulted all EBSCOhost databases. EBSCOhost entails Academic Search Elite, Business Source Premier, GreenFILE, Library, Information Science & Technology Abstracts (LISTA), PsycArticles, Psychology and Behavioral Sciences Collection, PsycINFO, and Regional Business News databases.

Table 1.1. *Search terms*

Search terms for review articles on predicting non-completion in higher education	
1.	Context: "university" OR "college" OR "higher education" OR "distance education" OR "online education" OR "online course" OR "adult education" AND
2.	Target group: "learner" OR "student" OR "undergraduate" AND
3.	Outcome measure: "stud* success" OR "stud* performance" OR "complet*" OR "drop* out" OR "persist*" OR "attrition" OR "achiev*" OR "progress*"
Search terms for intervention studies to raise completion rates in higher education	
1.	Context: "higher education" OR "university" OR "distance" AND
2.	Outcome measure: "dropout" OR "non-completion" AND
3.	Intervention studies: "intervention" OR "prevention" OR "program"
Complemented by additional search terms in a second literature search:	
4.	"matching" OR "selection" OR "study choice" OR "study decision"

Predictors. The search terms for the predictors of non-completion in higher education are presented in Table 1.1. This search was executed between March and April 2018. To find review studies on predictors of non-completion we defined search terms concerning context, target group and outcome measure and applied them for "all text." Several inclusion criteria were identified concerning review articles on predicting non-completion. These are presented in Table 1.2. The initial database search resulted in 929 articles. Duplicates were removed manually, resulting in 902 unique articles. These articles were screened based on the inclusion criteria by title and abstract, and if

necessary and available, whole text. If the whole text was required but not available, it was requested by contacting the authors. After full-text reading, eight review articles were included. A considerable number of articles was excluded in this step, because the outcome measure of completion was related to a medical field, such as treatment completion for drug abuse. Two articles were already at our disposal before database search, and met the inclusion criteria. These additional articles were included, resulting in a total of ten articles. This selection process is presented in Figure 1.1.

Table 1.2. *Inclusion criteria*

Inclusion criteria for review studies on predictors of non-completion in higher education	
1.	The article is peer-reviewed and published in an academic journal
2.	This article is a review or meta-analysis
3.	The outcome variable is non-completion or related (persistence, retention, attrition, dropout)
4.	The article is written in English, Dutch or Flemish
5.	The target group is in higher (online) education
6.	The target group is not a highly specific target group (e.g. minorities, students with a disability)
7.	The independent variables are within the scope of our review
Inclusion criteria for intervention studies to raise completion rates in higher education	
1.	The article is peer-reviewed and published in an academic journal
2.	The outcome variable is non-completion or related (persistence, retention, attrition, dropout)
3.	The article is written in English, Dutch or Flemish
4.	The study entails an investigation of an intervention with the purpose to increase completion rates in higher (online) education
5.	The target group is in higher (online) education
6.	The target group is not a highly specific target group (e.g. minorities, students with a disability)
7.	The intervention is within the scope of our review (e.g. interventions originate from the institution itself and not from for instance, governance funding of students etc.)
8.	The article is published in or after 2000

Interventions. The search terms for intervention studies designed to raise completion rates in higher education are presented in Table 1.1. To find relevant intervention studies the same search terms as mentioned above supplemented with "intervene*" or "prevent*" or "program" were applied. The most relevant hits were found using the search terms presented in Table 1.1. The database search for this part of the literature review was executed between May and June 2018 and later extended with a complementary search when it appeared that the results based on the initial search terms did not yield any interventions prior to student enrolment. For the intervention studies, we also defined some selection criteria, as presented in Table 1.2. Results of the database search were refined using relevant major heading and subject tags in EBSCOhost. The initial search and complementary search together resulted in 162 unique articles (134 from the initial search, 28 from the complementary search). These articles were screened based on the inclusion criteria, first on title and abstract. Again, if the title and abstract did not provide sufficient information the full text of the article was screened. After screening on title and abstract, there were 21 articles left (16 from the initial search, 5 from the complementary search). The screening of full text articles resulted in eight remaining articles (6 from the initial search, 2 from the complementary search). All articles

selected for whole text screening were discussed with all members of the research team, until consensus was reached. By applying the snowballing technique (i.e. checking the references of the included articles to find more relevant articles), eight additional articles were included (5 initial, 3 complementary search). Thus, after the first literature search for intervention studies we included sixteen articles. The selection process is presented in Figure 1.2.

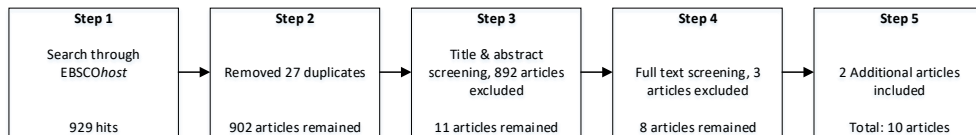


Figure 1.1. *Flowchart of the paper selection process for review studies on factors predicting non-completion in higher education*

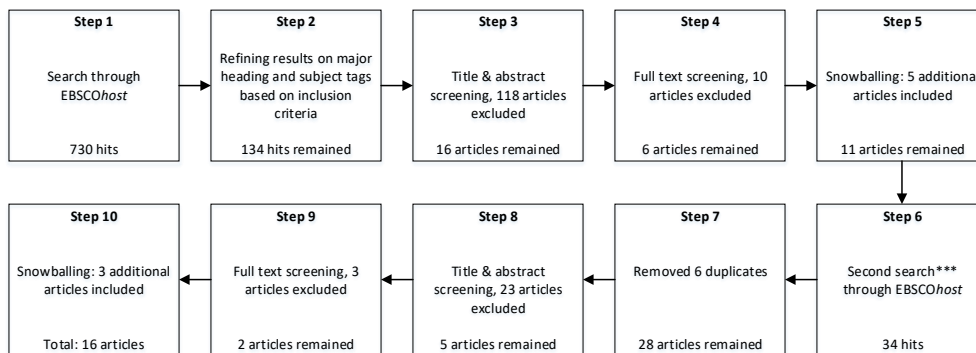


Figure 1.2. *Flowchart of the paper selection process on intervention studies*

1.2.2. Data generation and synthesis

Predictors. To obtain the results of the review studies on factors predicting non-completion, a data abstraction form was created, of which the components are presented in Table 3. In addition, the following data was extracted into a second form to evaluate the quality of the review studies: whether the databases and search terms as well as inclusion and exclusion criteria were given, the number of studies included, whether definitions and operationalization of (in)dependent variables were provided and whether the authors discussed the generalizability of both their review results and the individual studies they included. Two researchers independently summarized the articles according to these two forms, after which they discussed differences with each other and the other members of the research team until agreement was met. The results to evaluate the quality of the review studies is presented in Appendix A. As a vehicle to present our findings on predictors of non-completion consistently and concisely, we have chosen the generic model by Cross (1981). This model differentiates between three categories of variables related to student participation in higher education. First, dispositional factors are defined as individual factors, internal to the student, which may inhibit students'

participation in higher education. Carroll, et al. (2009) refer to beliefs, values, attitudes and perceptions in defining dispositional factors. Second, situational factors are defined as factors related to the circumstances in students' particular lives, for instance, employment and family commitments. Third, institutional factors are defined as “factors outside of the student's control, but those factors resulting from procedures, policies and structures of the educational institution that are related to students' participation in higher (online) education” (Carroll et al., 2009, p. 199). The simple distinction between these three categories makes the model very suitable as an initial framework to organise the wide variety of results from different studies. Considering our purposes, however, it became clear early in the process of reviewing that the model would benefit from a small extension, namely a subdivision of the category of dispositional factors into dispositional cognitive factors (i.e. ability or relevant knowledge, skills and experiences) and dispositional non-cognitive factors (i.e. affective and attitudinal factors). In addition, a category of demographic factors was added to the model. Figure 1.3. presents the full classification framework used. Two researchers independently categorized the results and uncertainties or differences between the categorization of the two researchers were discussed with the other members of the research team until consensus was reached.

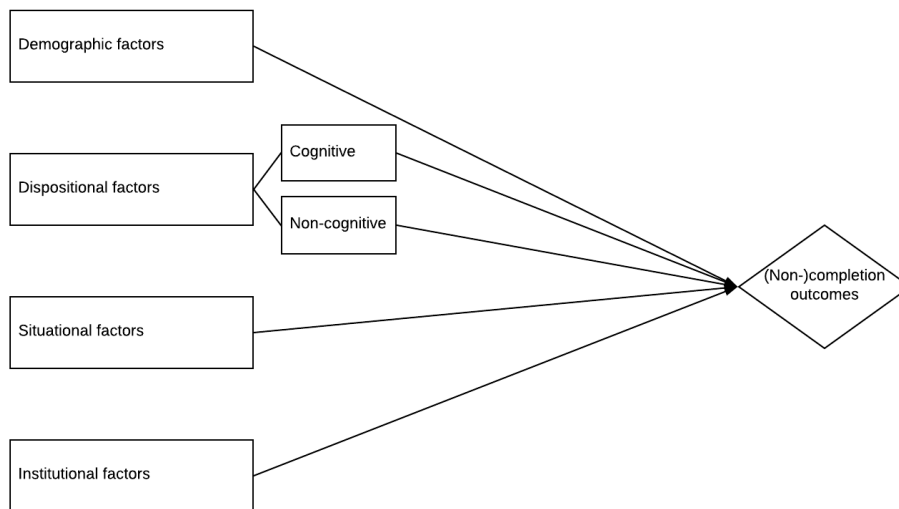


Figure 1.3. *Classification framework (Adapted from Carroll et al., 2009)*

Interventions. For the review on intervention studies, the same data extraction procedure was followed for partly different data, as presented in Table 1.3. To answer the second research question and identify the characteristics of effective interventions for raising completion rates we focused on the following characteristics:

- Intervention approach or strategy (e.g. mentoring, remedial teaching).
- Targeted factors (from the categories from the classification framework, see Figure 1.3.).
- Mode (online intervention, face-to-face intervention or a combination).

- Context (traditional higher education, online higher education or both).
- Duration of the intervention.
- Effect (whether the intervention raised completion rates significantly, effect size(s), and differences in completion rates between groups or cohorts).
- Cost effectiveness.

Interventions were categorized based on similarity of the treatment as coaching or remedial teaching, peer mentoring, motivational contact, academic dismissal policies or interventions on instruction, to present the results in an organized manner. With regard to the quality of the intervention studies, we classified the sample size, whether the sampling method was discussed, whether the intervention method and decision for a target factor were theoretically underpinned and whether authors discussed generalizability of their results, and possible threat to internal validity. The results with regard to the quality of the intervention studies are presented in Appendix B.

Table 1.3. *Data extraction components*

Data extraction components for review articles on predicting non-completion in higher education	
1.	Reference
2.	Educational context
3.	Outcome measure (definition and operationalization)
4.	Independent measure(s) (definition and operationalization)
5.	Results
6.	Conclusion
Data extraction components for intervention studies to raise completion rates in higher education	
1.	Reference
2.	Research question
3.	Purpose of the study
4.	Sample (size)
5.	Factors manipulated or targeted at by the intervention (e.g. academic self-efficacy or motivation)
6.	Description of the intervention
7.	Duration of the intervention
8.	Theoretical underpinning of the intervention instrument and the target factor
9.	Outcome measure related to non-completion
10.	Results
11.	Conclusion

1.3. Results

1.3.1. Predictors of non-completion

Quality appraisal. Before describing the results, we discuss the quality of the review studies included in the first part of the review. We also scored the included articles on the quality criteria discussed in section 1.2.2., for which we refer to Appendix A. We have found 10 review studies (see Table 1.4.), of which only two were meta-analyses that applied certain quality criteria (e.g. effect sizes), as a threshold for including studies in their review (Fong, et al., 2017; Robbins et al., 2004). The other studies provide a more narrative overview, or provide a systematic overview without reporting quantitative results (Bowles

& Brindle, 2017; Credé & Niehorster, 2012; Lee & Choi, 2011; O'Neill et al., 2011; Pascarella, 1980; Riggert et al., 2006; Trapmann et al., 2007; Van Rooij et al., 2018). The number of studies/ articles taken into account for individual factors in the review studies ranged from 6 (for six factors in Robbins et al., 2004) to 36 (for one factor in Robbins et al., 2004). Nine out of ten review studies discussed which databases were used to find relevant articles, and six of them defined and reported search terms. Nine review studies presented in- or exclusion criteria used in screening articles. Important to take into account when interpreting the results presented in the next paragraph, is that there were considerable differences in operationalization and definition of the same variables included in different review studies (e.g. motivation as defined and measured by Robbins et al., 2004 and Fong et al., 2017). In some review studies, specific definitions and operationalization used in the individual studies they have included were not discussed. In terms of generalizability, some review studies focused on predictors of non-completion in a specific country (Van Rooij et al., 2018) or a specific study program (e.g. O'Neill et al., 2011). Eight out of ten review studies discussed generalizability of their findings. With respect to generalizability it is important to note that two review studies (although they discussed generalizability of their results) reported significant results only, leaving it unclear to what extent the individual studies included in their review also investigated the predictive value of other variables without significant results. The results of these two review studies may be generalizable, but they leave out important information and in doing so have a limited contribution to obtaining a comprehensive picture. Based on our assessment of the quality of the review studies, we decided to exclude some predictors discussed in these studies from further analyses, because their definition and operationalization appeared not sufficiently distinct from the independent (outcome) variables (e.g. persistence, dropout). For instance, we excluded academic struggling, operationalized as the amount of failed science tests in the first year of higher education, grade point average in the first year of higher education and decelerated curriculum status (O'Neill et al., 2011), academic momentum and academic success (Bowles & Brindle, 2017), and current grade point average (Lee & Choi, 2011). This was, to us, not enough reason to exclude these review studies fully from analyses, though this explains why not all variables from all review studies will be discussed in the results section. Next, the results of the review studies on predictors of non-completion in higher education will be described, organized in the categories as explained in section 1.2.2. These results are presented in Appendix C and an overall synthesis of the results is presented in Figure 1.4.

Table 1.4. *Overview of the included articles and the corresponding categories from the theoretical framework on predictors of non-completion in higher education*

*	Reference	Categories
1.	Pascarella (1980)	Institutional
2.	Robbins et al. (2004)	Demographic, Dispositional cognitive, Dispositional non-cognitive, Institutional
3.	Riggert et al. (2006)	Situational
4.	Trapmann et al. (2007)	Dispositional non-cognitive
5.	Lee & Choi (2011)	Dispositional cognitive, Dispositional non-cognitive, Situational, Institutional
6.	O'Neill et al. (2011)	Demographic, Dispositional cognitive, Dispositional non-cognitive, Institutional
7.	Credé & Niehorster (2012)	Dispositional non-cognitive
8.	Bowles & Brindle (2017)	Demographic, Dispositional cognitive, Dispositional non-cognitive, Situational, Institutional
9.	Fong et al. (2017)	Dispositional non-cognitive
10.	Van Rooij et al. (2018)	Demographic, Dispositional cognitive, Dispositional non-cognitive

Note. *These numbers are also used to refer to the articles in Appendix C.

1.3.2. Results on predictors of non-completion in the classification categories

Demographic variables. Four review studies (of which one meta-analysis) focused on demographic factors in relation to non-completion. All four studies focused on socioeconomic status, for which inconsistent results were found in relation to non-completion outcomes (Bowles & Brindle, 2017; O'Neill et al., 2011; Robbins et al., 2004; Van Rooij et al., 2018). Age, gender, and parents' education were all investigated in two review studies, and for all three factors, inconsistent results were found in individual studies (see O'Neill et al., 2011 for age, gender and parents' education; Bowles & Brindle, 2017 for age and parents' education; Van Rooij et al., 2018 for gender). Consistent results were found for the link between ethnicity and student dropout, though only investigated in one of the included review studies (O'Neill et al., 2011). All four studies included in that review by O'Neill et al. (2011) indicated no significant relation between ethnicity and drop-out.

Dispositional cognitive variables. Six review studies (of which two meta-analyses) included dispositional cognitive variables. One of the most consistent results is found for entry qualifications, like high school grade point average, and scores on pre-entry tests (i.e. in American higher education context, ACT or SAT scores). These factors showed to be significantly positively related to persistence outcomes (Lee & Choi, 2011; O'Neill et al., 2011; Robbins et al., 2004; Van Rooij et al., 2018). Five out of six review studies included learning or study strategy factors. Out of these five, four report a significant relation with non-completion (significant in Robbins et al., 2004; Lee & Choi, 2011; Bowles & Brindle, 2017; Van Rooij et al., 2018; not significant in Fong et al., 2017). The meta-analysis by Robbins et al. (2004) reports an estimated true correlation between academic-related skills and retention of 0.366. Important to note with respect to learning or study strategy factors is the difference in definition and operationalization within and between different review studies. Two out of six review studies focused on preparedness (Bowles & Brindle, 2017; Van Rooij et al., 2018), which was not a factor of interest in the other four review

studies. Inconsistent results between and within review studies were reported with respect to the link between this factor and non-completion outcomes. Factors investigated only in singular review studies were: number of online courses completed previously, experience in relevant field, involvement in professional activities, computer skills (Lee & Choi, 2011), and intelligence (Van Rooij et al., 2018). The factors investigated by Lee and Choi (2011) were all found to be negatively related to online course dropout. Intelligence was not found to be significantly related to persistence by Van Rooij et al. (2018), however, this was based on only one scientific study.

Dispositional non-cognitive variables. A large number of studies focused on dispositional non-cognitive factors. In total, eight review studies (of which two meta-analyses) focused on variables within this category (Bowles & Brindle, 2017; Credé & Niehorster, 2012; Fong et al., 2017; Lee & Choi, 2011; O'Neill et al., 2011; Robbins et al., 2004; Trapmann et al., 2007; Van Rooij et al., 2018). Five review studies included motivational factors, and investigated the relationship with non-completion outcomes (Bowles & Brindle, 2017; Fong et al., 2017; Lee & Choi, 2011; Robbins et al., 2004; Van Rooij et al., 2018). Four of them found positive significant relationships for motivational factors and persistence or retention outcomes (Bowles & Brindle, 2017; Fong et al., 2017; Lee & Choi, 2011; Van Rooij et al., 2018). Fong et al. (2017) reported a significant correlation of 0.150 in their meta-analysis. However, Robbins et al. (2004) reported a non-significant estimated true correlation of only 0.066. In addition, intrinsic motivation, as investigated by Van Rooij et al. (2018) was not found to be significantly related to retention in the majority of the studies they reviewed (non-significant in four studies, positively significant in two studies). Extrinsic motivation in their review study was consistently not related to persistence. In two out of the three studies, they have included 'study motivation', which was positively related to persistence. Lack of motivation was negatively related to persistence in two out of two studies included by Van Rooij et al. (2018). Differences in definition and operationalization of motivational factors within and between review studies complicate an accurate evaluation of these contradictory results.

Four review studies investigated self-efficacy (Robbins et al., 2004; Bowles & Brindle, 2017; Fong et al., 2017; Van Rooij et al., 2018), and reported consistent positive relationships between self-efficacy and persistence or retention outcomes. Robbins et al. (2004) found an estimated true correlation between self-efficacy and retention of 0.359, while Fong et al. (2017) reported a correlation between self-perceptions (including self-efficacy) and persistence of 0.100. Robbins et al. (2004) found no significant relationship between general self-concept and retention. Factors investigated in three review studies and resulting in consistent results were: goals and intentions (Robbins et al., 2004; Bowles & Brindle, 2017; Lee & Choi, 2011), institutional or college adjustment (Robbins et al., 2004; Credé & Niehorster, 2012; Van Rooij et al., 2018), and personality characteristics (Trapmann et al., 2007; Bowles & Brindle, 2017; Van Rooij et al., 2018). Goals and intentions were significantly positively related to retention or persistence outcomes (Robbins et al., 2004; Bowles & Brindle, 2017; Lee & Choi, 2011). Robbins and colleagues reported an estimated true correlation of 0.340 between academic goals and retention. Three review studies investigated the predictive value of institutional or college adjustment factors. These

factors refer to the extent to which a student has adapted to academic demands, which is defined by a student's attitude toward the study program or course, their engagement with the study material and the adequacy of their efforts in studying (Credé & Niehorster, 2012). These factors are thus clustered in the category of dispositional non-cognitive factors (and not to institutional factors), because they refer to processes inherent to the student, and not the institute. Institutional or college adjustment factors were significantly positively related to retention or persistence outcomes in all three studies investigating this link (Robbins et al., 2004; Credé & Niehorster, 2012; Van Rooij et al., 2018). Robbins et al. (2004) reported an estimated true correlation of 0.206 for this link, Credé and Niehorster (2012) reported an estimated true correlation of 0.230 for this relationship. Moreover, Credé and Niehorster (2012) reported effect sizes of subscales of institutional adjustment, in which the largest estimated true correlation was found between institutional attachment and retention of 0.290, followed by the predictive value of social adjustment (true score correlation = 0.250), academic adjustment (true score correlation = 0.190) and personal-emotional adjustment (true score correlation = 0.130). Inconclusive results between review studies were found with respect to the relation between personality characteristics and non-completion outcomes (Trapmann et al., 2007; Bowles & Brindle, 2017; Van Rooij et al., 2018). Attributions were examined as a predictor of non-completion in two of the review studies, for which different results were found (significantly related to non-completion in Lee & Choi, 2011; no significant results in Fong et al., 2017). For results other dispositional non-cognitive factors we refer to Appendix C, as they were investigated in only one of the included review studies, for instance anxiety, which was not significantly related to completion outcomes (Fong et al., 2017) and difficulty juggling commitments, which was negatively related to completion outcomes (Bowles & Brindle, 2017).

Situational variables. Three of the included review studies investigated the relationship between situational variables and non-completion outcomes. The relationship between employment factors and non-completion outcomes was investigated in all of these three review studies (Bowles & Brindle, 2017; Lee & Choi, 2011; Riggert et al., 2006). While Lee and Choi (2011) and Bowles and Brindle (2017) reported a straightforward positive relationship between employment pressures or commitments and student dropout, Riggert et al. (2006) reported a more complex relationship between employment and completion outcomes. This latter review indicates that 1–15 employment hours) might be beneficial for completion rates as compared to no employment commitment at all. Financial aid or scholarship (Bowles & Brindle, 2017; Lee & Choi, 2011), and supportive social networks (Bowles & Brindle, 2017; Lee & Choi, 2011) were investigated in two out of three review studies. Financial aid or attainment of a scholarship are consistently positively related to completion outcomes, as are supportive social networks or emotional support (Bowles & Brindle, 2017; Lee & Choi, 2011). Other factors were investigated in only one review study (see Appendix C). For instance, family responsibilities or pressures (e.g. from controlling parents) relate negatively to completion outcomes (Bowles & Brindle, 2017).

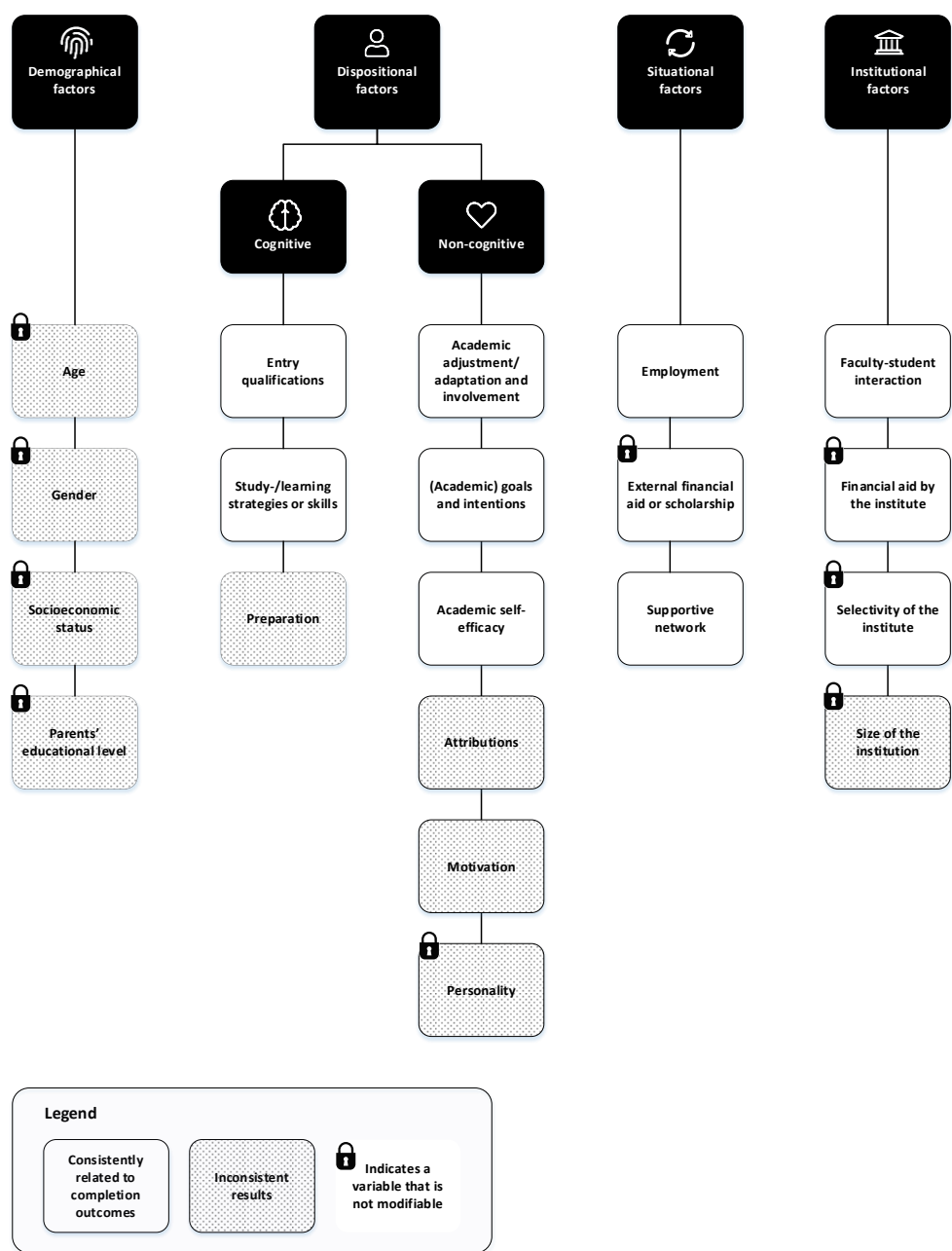
Institutional variables. Five review studies (of which one meta-analysis) investigated the relationship between institutional variables and non-completion outcomes (Bowles & Brindle, 2017; Lee & Choi, 2011; O'Neill et al., 2011; Pascarella, 1980; Robbins et al., 2004). Three of these investigated the relationship or interaction between faculty (staff) and students, reporting significant positive relations with persistence. In only one out of seven individual studies included by Pascarella (1980) no significant relationship was found. Financial support by the institute, size of the institute, and selectivity of the institute, were investigated by two review studies (Bowles & Brindle, 2017; Robbins et al., 2004). Both studies report a significant positive relationship for financial support (estimated true correlation of 0.188 in Robbins et al., 2004). For size of the institute, an estimated true correlation of -0.010 was reported by Robbins et al. (2004), which was not significant. A significantly negative relationship was found between size of the institute and retention rates by Bowles and Brindle (2017). For institution selectivity (i.e. the extent to which educational institutions set a standard for selecting new students) a significant positive link with retention outcomes was reported by Robbins et al. (2004) (estimated true correlation = 0.238) and Bowles & Brindle (2017). All other factors in this category were investigated in one review study only, for which we refer to Appendix C. For instance, curriculum type, which is investigated by O'Neill et al. (2011), reporting higher student dropout in traditional curriculum type, as compared to a problem-based learning curriculum type.

1.3.3. Synthesis of results on predictors of non-completion

One of the aims of this review study was to create an overview of (modifiable) variables that are related to non-completion in higher education. In Figure 1.4., we present an overview of the variables related to non-completion, based on the results of this literature review and categorized according to the model presented in Figure 1.3. We indicated whether factors are modifiable (i.e. changeable or to be advised on) by putting a lock on those variables that are not modifiable. We did not take into account variables investigated by only one of the included review. In this Figure, variables are presented in alphabetical order (per category of the theoretical framework).

All in all, modifiable consistent predictors of non-completion in higher education are study- or learning strategies, academic self-efficacy, (academic) goals and intentions, institutional or college adjustment, employment, supportive network and faculty-student interaction. For these factors there were three review studies providing effect size by means of estimated true correlations. The most effective modifiable consistent predictors for non-completion based on these review studies seem to be study-/learning strategies or skills (estimated true correlation of 0.366, see Robbins et al., 2004), academic goals and intentions (estimated true correlation of 0.340, see Robbins et al., 2004), academic adjustment or adaptation and involvement (estimated true correlations of 0.206–0.230, see Robbins et al., 2004 and Credé & Niehorster, 2012), and academic self-efficacy (estimated true correlation of 0.359, see Robbins et al., 2004). We need to take into account some points in interpreting these results. Some factors that might be modifiable were not investigated in a thorough number of review studies (e.g. computer skills in the category dispositional cognitive factors). There are also consistent predictors of non-

completion in higher education that do not seem modifiable, but maybe are. Entry qualifications in the category of dispositional cognitive factors might be such a factor. Some entry qualifications cannot be changed, of course (e.g. grade point average in high school). However, other entry qualifications, mathematical skills for instance, might be subject to interventions in which this factor is tested and remedial teaching is provided if necessary. Employment itself cannot be changed by interventions implemented by educational institutions, however the amount of employment hours also gives an indication about the amount of hours students can spend on their studies, on which students can be advised by educational institutions. Therefore, we did not put a lock on the employment factor. Important to note is that due to a lack of comparability and effect sizes, the results on modifiable predictors of non-completion are still rather inconclusive. Especially in the category of dispositional non-cognitive factors there is a lack of comparability, because overlapping constructs are operationalized differently (e.g. academic study skills and learning strategies) or the same operationalization is used for a slightly different construct (e.g. self-esteem questionnaires used to measure self-concept) and in the majority of the review studies definitions or operationalization of constructs are not provided. Finally, with respect to generalizability of these results, only two of the review studies concerned a higher online educational context, which means that drawing conclusions on predictors of non-completion in this context should be done with caution.



Note: This Figure is designed in Microsoft Visio©

Figure 1.4. Overview of variables related to non-completion in higher education

1.3.4. Intervention studies

In the results section of the intervention studies, the interventions and the corresponding results with respect to completion rates are described first, grouped in different categories of interventions (see Table 1.5.), in chronological order. After that, in section 1.3.6., an overview will be presented of the characteristics of effective and efficient interventions, in line with our second research question. The characteristics we focus on are based on the data extraction components and were discussed in section 1.2.2.

Table 1.5. *Overview of the included articles on interventions to raise completion rates in higher education and the corresponding category of interventions*

*	Reference	Categories
1.	Wang & Grimes (2000)	Coaching/Remedial Teaching
2.	Chyung (2001)	Intervention on Instruction
3.	Pagan & Edwards-Wilson (2002)	Peer Mentoring
4.	Ruthig et al. (2004)	Coaching/Remedial Teaching
5.	Salinitri (2005)	Peer Mentoring
6.	Huett et al. (2008)	Motivational Contact
7.	Simpson (2008)	Motivational Contact
8.	Larose et al. (2011)	Peer Mentoring
9.	Martorell & McFarlin (2011)	Coaching/Remedial Teaching
10.	Stegers-Jager et al. (2011)**	Academic Dismissal Policy
11.	Bettinger & Baker (2014)	Coaching/Remedial Teaching
12.	De Paola & Scoppa (2014)	Coaching/Remedial Teaching
13.	Patterson et al. (2014)	Coaching/Remedial Teaching
14.	Arnold (2015)**	Academic Dismissal Policy
15.	Inkelaar & Simpson (2015)	Motivational contact
16.	Sneyers & De Witte (2017)**	Academic Dismissal Policy

Note. *These numbers are also used to refer to the articles in Appendix D. **These articles concern the same intervention for overlapping data sets. Article 14 is about Dutch university samples from 2002-2007, Article 16 is about Dutch higher education samples (including university samples) from 2003-2004 and 2008-2009 and Article 10 is about a specific single Dutch university sample from 2003-2004 and 2005-2006.

Quality appraisal. Before elaborating on the results of the intervention studies, we will as in part one, first discuss the quality of the intervention studies included in this literature review. The included articles are scored on these quality criteria in Appendix B. As presented in Table 1.5., 16 intervention studies have been included in the present literature review. Four of these intervention studies were carried out (at least partly) in the context of higher online education (Chyung, 2001; Huett et al., 2008; Inkelaar & Simpson, 2015; Simpson, 2008). The total number of participants in these intervention studies ranged from 12 (Chyung, 2001) to 255878 (Martorell & McFarlin, 2011). Six of the interventions investigated were (at least partly) online interventions (Bettinger & Baker, 2014; Chyung, 2001; Huett et al., 2008; Inkelaar & Simpson, 2015; Ruthig et al., 2004; Simpson, 2008). Interventions lasted from a minimum of one informal session (Ruthig et al., 2004) until one year (Arnold, 2015; Bettinger & Baker, 2014; Larose et al., 2011; Salinitri, 2005; Sneyers & De Witte, 2017; Stegers-Jager et al., 2011), though not all intervention studies gave details regarding the duration of the intervention. In terms of generalizability, there are several points that require attention. Some of the results in these intervention studies are based on rather small sample sizes (Chyung, 2001; Salinitri, 2005; Simpson, 2008), some of the interventions are evaluated for rather specific target

groups, although most of the underlying mechanisms in these interventions seem generalizable to other target groups as well. For example, the intervention by De Paola and Scoppa (2014) was investigated in the Italian educational context, which is (in the explanation the authors provided) comparable to the traditional Dutch higher educational context. In contrast, some parts of the intervention by Chyung (2001) are inherent to the educational context in which the intervention is investigated (specifically for students enrolled in the 'Instructional and Performance Technology' program), resulting in decreased generalizability of the intervention to other educational contexts. In ten of the included intervention studies generalizability of the results was discussed. In terms of threat to internal validity, also multiple points need to be stressed. For instance, in a majority of the intervention studies there has been no manipulation check, to analyse whether the factor that was aimed to be modified (e.g. motivation), actually changed by the intervention (Bettinger & Baker, 2014; De Paola & Scoppa, 2014; Inkelaar & Simpson, 2015; Pagan & Edwards-Wilson, 2002; Patterson et al., 2014; Ruthig et al., 2004; Salinitri, 2005; Simpson, 2008). In addition, in some intervention studies actually multiple interventions are evaluated at once, which makes it hard to interpret the results on effectiveness of the intervention characteristics (Chyung, 2001; Huett et al., 2008; Wang & Grimes, 2000). In some intervention studies a control group was included, however, in some cases this entailed a passive control group, which means that results on effectiveness of the intervention might also be due to the fact that the experimental group underwent at least some procedure, independent from what the actual procedure entailed (e.g. Inkelaar & Simpson, 2015; Larose et al., 2011). Additionally, in some intervention studies there might have been a self-selection bias (e.g. based on first come, first served principle for remedial teaching or voluntary basis) (e.g. Patterson et al., 2014; Ruthig et al., 2004). In eleven of the intervention studies, there was attention for possible threats to internal validity by either addressing them in discussing their findings or even taking measures to prevent threats to internal validity. The results of the intervention studies are presented in Appendix D.

1.3.5. Results on intervention studies in the intervention categories

Coaching and remedial teaching. In this category of interventions, we discuss results of interventions in which students received some sort of coaching/mentoring or remedial teaching by professional teachers, trainers or coaches. Wang and Grimes (2000) evaluated the Access Plus Program in traditional higher education. This program involved multiple offers for freshmen in college, for instance an advising program, a seminar course, interest groups, and remedial teaching for English and mathematics. The duration of this intervention and number of participants included in the study were not specified. The Access Plus Program aimed at improving academic motivation, social motivation, general coping skills and receptivity to institutional support, which were all measured prior to the start of the intervention by the College Student Inventory. However, no post-measurement was carried out. It was reported that after this intervention there was a 10% increase for freshmen to sophomore (the second) year.

Ruthig et al. (2004) investigated an optimism and attributional retraining program in the context of traditional higher education. This program consisted of an

informal session, which was executed differently in three groups. The information of interest in this informal session was presented by either a videotape of 8 min, the videotape followed by a 20-min group discussion or a handout only. Theories underlying the intervention were explained, for instance, unrealistic optimism and attributional theories). In this attributional retraining positive effects of effort attributions (i.e. "I failed this test, because I did not put enough effort in studying the course material") on college performance were emphasized, in contrast to ability attributions (e.g. "I failed this test because I am not smart enough"). Dispositional optimism was measured prior to the intervention. It was concluded that this intervention decreased voluntary course withdrawal significantly, but only for high optimism students who received attributional retraining.

Martorell and McFarlin (2011) examined the effect of developmental education (as part of the broader Texas Academic Skills Program) on mathematics, reading and writing in 2-year and 4-year study programs in traditional higher education. This was a face-to-face intervention, of which the duration and theoretical underpinning was not specified. This intervention was targeted at basic skills in a number of courses, such as mathematics and language skills. Assignment to the remedial teaching courses was based on diagnostic tests. No detailed description was provided with respect to the remedial teaching itself. Significant results were found in the 2-year study program context only, and showed that fulfilling these remedial courses, in contrast to what expected, lowered the probability of completing at least one year in college by 6%, only when controlling for baseline covariates, such as age, ethnicity, and academic year of enrolment.

Bettinger and Baker (2014), in a randomized experiment, researched the effectiveness of individualized student coaching provided to students in public, private and proprietary universities by a student coaching service called InsideTrack. This intervention was based on three barriers for completion in higher education, identified in prior research: the lack of appropriate information, the lack of students' academic preparation and the lack of integration in the university community. Within the service of InsideTrack (a for-profit provider of coaching services), students are matched to coaches. Coaches contact students on a regular basis, by phone calls, email, text messages and social networking sites, to provide help and support in the beginning of the students' college careers. Coaches working for InsideTrack are hired through a very rigorous application procedure. Phone calls are recorded and coaches receive feedback on the content and tone of their phone calls with students. InsideTrack aims for a 20% institution-specific and 80% general content ratio in the contact between coach and student and in some cases coaches have access to study materials. After 6 and 12 months of this intervention, the persistence rate for coached students was significantly higher than for students who did not receive InsideTrack's coaching. After 18 and 24 months, the difference in persistence rates between the coached and control students is still significant at the 1% level, even though the coaching lasted only 12 months. The results do not change when controlling for covariates like ACT/SAT scores, age, high school GPA or scholarship.

De Paola and Scoppa (2014), like Martorell and McFarlin (2011), investigated the effectiveness of mathematics and language skills remedial courses in the context of

traditional higher education. This face-to-face intervention lasted two months, and entailed 160 h of remedial teaching. Remedial teaching was implemented at the beginning of the academic year and students were assigned based on their performance on a placement test. Although participation was strongly recommended, it was not compulsory. No detailed description of the remedial teaching was provided. A decrease in non-completion probability between 6 and 13.5% was demonstrated for students attending 100 h of remedial courses, which was statistically significant at the 10% level.

The last intervention study in this category, by Patterson et al. (2014) investigated a face-to-face self-regulated learning course for students in traditional higher education. The duration of this intervention was not specified. Within this self-regulated learning course, there was a focus on critical thinking skills and an effort was made to guide students in taking control of their academic lives, aimed at improving students' autonomy. Four self-regulated learning strategies were included: discovering questions pertaining to a course and the methodology for answering them, cognitively engaging with material, identifying teachers' goals and working to meet them, and monitoring one's own comprehension. In addition, students learned techniques to fulfil these strategies, like active reading, creating concept elaborations and developing mock exams. The self-regulated learning course was a 3-credit elective that any undergraduate student could take. This course entailed 50-min lectures twice a week, and weekly meetings in which students showed and discussed their application of self-regulated learning strategies, on which peer monitors provided feedback. Results showed to be significant at the 1% level and indicated that students who completed the self-regulated learning course in the first year were approximately twice as likely to be enrolled in the second year. This effect lasted until the fifth year of college.

Peer mentoring. In this category, we discuss interventions comparable to the previous category, as they are also on coaching and mentoring. However, in this category we specifically discuss coaching and mentoring provided by peers (trained to serve as a coach/mentor), in contrast to professional teachers, trainers or coaches. Pagan and Edwards-Wilson (2002) examined the effectiveness of a mentoring program for at-risk students (students on academic warning or probation). The mentoring program lasted for one year and was targeted at improving completion rates through improvement of students' academic and interpersonal skills. These factors were, however, not measured in the intervention study. Mentors were selected for an interview from a list of students with high GPA scores and who volunteered to serve as mentors, and eight of them were hired eventually as a mentor. Mentors attended required training sessions, staff meetings and weekly supervision and they received written materials about the theories underlying the mentoring program discussed in the training and meetings. The mentoring program itself consisted of an orientation meeting in which contracts and the goals and responsibilities were discussed. After this meeting, mentors contacted the mentees via email and personal note cards including information to make a face-to-face appointment. Eventually, if mentees did not make a face-to-face appointment, they were contacted by phone. Overall, mentors met with their mentees at least twice, had contact via email and held phone conversations. During the meetings, a specific protocol was followed in which study skills, financial aid, and personal issues were discussed. Statistical analyses of

effects were carried out only in relation to GPA of the mentees. Descriptive results reported on non-completion showed that after the mentoring program the status of the 53 students initially on academic warning or probation changed to: 23 students retained in good academic standing, 3 retained on warning, 6 retained on probation and 21 students were academically dismissed.

Salinitri (2005) investigated the effects of a mentoring program in traditional higher education. This mentoring program lasted for one year and was targeted at social and academic integration. In this mentoring program, teaching candidates were mentors for first-year students. It was aimed that this mentoring intervention would build networking, skills in self-concept and strengthen the goals of first-year students. The mentors were enrolled in a course in which practices of mentoring, advising and social learning were discussed. Mentors were instructed to journal their activities of the mentor meetings and to write reflective summaries of their experiences. Mentees were asked to assess the mentors' skills by means of the Mentor Assessment Survey. This intervention was executed and evaluated twice. Enrolment as a mentee in the mentoring condition was on voluntary basis. Results showed to be significant at the 1% level, in which in the first run of the intervention a retention rate of 88.5% was found in the group who received mentoring, as compared to 57.1% in the control group. In the second run of the intervention, a retention rate of 71.4% was found for the group who received mentoring, as compared to 23.1% in the group who did not receive mentoring.

Larose et al. (2011) evaluated the effectiveness of a peer-mentoring program in traditional higher education, more specifically, a math, science and technology program. A socio-motivational mentoring model constituted the theoretical underpinning of the intervention, which explicitly targeted college adjustment, motivation and career decision. In this peer-mentoring program, there were bimonthly meetings between mentors and mentees. Mentors were selected based on previous experience, college performance, and their ability to deal with relationship issues. Mentors and mentees were matched as much as possible according to college, program, professional interests and gender. Mentors were trained in a two-day training seminar and guided by eight supervisors during the implementation of the intervention. Mentors were asked to complete a logbook about the meetings with their mentees. The effectiveness of the program was evaluated by a randomized pre-test/post-test control group design. Motivation, career decision profile and adjustment to college were measured prior and after the intervention by the Academic Motivation Scale, Career Decision Profile Inventory and Student Adaptation to College Questionnaire respectively. After the intervention, mentees showed significantly higher levels of motivation, institutional attachment, social adjustment and a more positive career decision profile, as compared to the students in the (passive) control group. Results demonstrated that this intervention raised completion rates significantly: 86% compared to 76% in the control group.

Motivational contact. In this category, intervention studies are discussed in which students received motivational support by means of e-mail messages, phone calls or letters. Huett et al. (2008) sent motivational emails and investigated the effect of these emails on withdrawal in both higher online education and traditional higher education. This intervention lasted one course or semester and was targeted at improving

completion rates through improvement in ARCS factors (i.e. attention, relevance, confidence and satisfaction), which were measured by the Course Interest Survey. The experimental groups were sent simple, mass-mailed motivational emails throughout the semester, entailing an enthusiastically written introduction (e.g. "I hope you are doing great"), goal reminders (e.g. "Don't forget the deadline for ..."), words of encouragement (e.g. "You can do it"), and multiple points of contact (e.g. "Do not hesitate to contact ..."). This intervention showed to be significant at the 5% level, but only in the online context.

Simpson (2008) also investigated the effect of motivational emails, supplemented by motivational telephone contact and letters, in a higher online educational context. This intervention lasted one course, and was based on a broad range of theories, among which ARCS factors, self-determination theory, and the strength approach. The content and procedures of the telephone and email contact were not further specified. It was discussed that motivational telephone contact only increased retention by around 5% and the combination of motivational emails, letters and telephone contact increased retention by around 25 percentage points.

Inkelaar and Simpson (2015) evaluated the effect of motivational emails only in higher online education, in an intervention that lasted approximately six months. The theoretical underpinning mentioned for this intervention was, like in the two studies discussed previously the ARCS factors, theories of self and positive psychology. Motivational emails were sent biweekly, comprised messages of around 400 words, were addressed personally to a student (instead of 'Dear student'), were signed by a person designated as 'University of London Learning Consultant' and were written in an informal friendly style containing suggestions about learning and overcoming learning problems. The emails were called 'Study Tips' and seventeen topics were addressed in a corresponding number of emails. For example, motivating yourself to learn, making lists, learning to concentrate on learning and exam tactics. A monitor showed that approximately 37.3% of the recipients on average opened the emails. This intervention appeared to be significant at the 10% level only, and an increase of 2.3 percentage points in retention was presented.

Academic dismissal policies. In this category, interventions are discussed in which there is a form of 'selection after enrolment', by means of academic dismissal (AD) policies. Important to keep in mind reading these results is that they were partly based on the same data. Stegers-Jager and colleagues (2011) evaluated an academic dismissal policy implemented in a specific context of medical education. Two AD cohorts were compared to two non-AD cohorts on several outcomes, among which dropout rates and year 1 curriculum completion. This intervention consists of two components. First, students were warned when they failed to meet set standards. In addition, students who were warned were offered academic support meetings on a voluntary basis. The results showed that there was a significant difference in dropout rate in terms of completing the first year curriculum (measured 2 years after enrolment). The effect size was 0.07.

Arnold (2015) examined the effectiveness of academic dismissal policies in Dutch (traditional) universities in cohorts from 2002 until 2007. In academic dismissal policies in the Netherlands a binding study advice is given, based on the number of study credit points obtained during the first year in university. Below a certain threshold of attained

study credits students receive a negative, binding study advice. Students who obtained the maximum amount of credits receive a positive advice and students in between the threshold and the maximum receive a conditional positive study advice. This means in most of the cases that these students have to obtain all first year credits before the end of the second year. In most institutes, students who received a negative binding study advice are provided support in their transition to another degree program. The function of these academic dismissal policies is twofold. On the one hand these policies have a selective function (i.e. "preventing students from spending too much time in pursuing a study for which they do not have the skills, talent or motivation", p. 1071). On the other hand, it has a referential function (i.e. "putting students in the right track in time", p. 1071). The results showed that overall the academic dismissal policies increased non-completion in the first year by an average of 6–7%. However, completion rates after four years improved by 5–9%. Overall, first year dropout rate for students in AD cohorts is 35.8%, compared to 27.9% for students in non-AD cohorts. These differences are significant at the 1% level.

Sneyers and De Witte (2017) also investigated academic dismissal policies in the Netherlands, for both research universities and universities of applied sciences (both traditional higher education), for cohorts from 2003 to 2004 and 2008–2009. Their results are in line with the results from Arnold (2015) and suggest that the implementation of an academic dismissal policy results in higher first year non-completion, but also a higher graduation rate (completion rate after four years). Significant at the 0.01% level, they showed that first-year non-completion will increase by 7.5% by implementation of an academic dismissal policy.

Interventions on instruction. In the last category of interventions, we discuss intervention studies focussing on the effect of changes in instruction and delivery method of education on completion rates. Chyung (2001) investigated the combined effect of diverse systematic instructional methods in online courses as an intervention to raise completion rates. In total, the study mentions 28 instructional methods linked to the ARCS constructs. For instance, class sizes were kept small (about 17 students), learners were provided with a technical training program, clearly stated weekly goals were provided, personal contact was made with each learner through a personal discussion area online or email, and multimedia materials were used in instruction. The intervention lasted one course or semester and the ARCS variables were measured prior and after the intervention. The questionnaires were filled by 12–20 participants, yet it was not specified on how many students the figures on retention were based. Results showed that before the intervention was implemented 44% of the students dropped out of the program by their third course. After the first cycle of implementation, this figure decreased to 22% and a further 15% in subsequent years.

1.3.6. Synthesis of characteristics of effective and efficient Interventions

The second aim of this review study was to gain insight in the characteristics of effective and efficient interventions to raise completion rates in higher education. In Figure 1.5., we present the effectiveness and characteristics (see section 1.2.2.) of all categories of interventions included in this literature. Even though cost-effectiveness might be an

important characteristic to take into consideration during design of an intervention, it is not taken into account in Figure 1.5., as in all categories of interventions information regarding this characteristic was lacking. All in all, interventions raising completion rates significantly are coaching or remedial teaching and peer mentoring, in which the differences between experimental and control groups or cohorts before and after implementation of interventions range from 6 to 54 percentage points. Academic dismissal policies decrease completion rates in the first year (an effect size of 0.07 reported by Stegers-Jager and colleagues, 2011), but by selection after student enrolment increasing completion rates in the long run (i.e. graduation rates after 4 years).

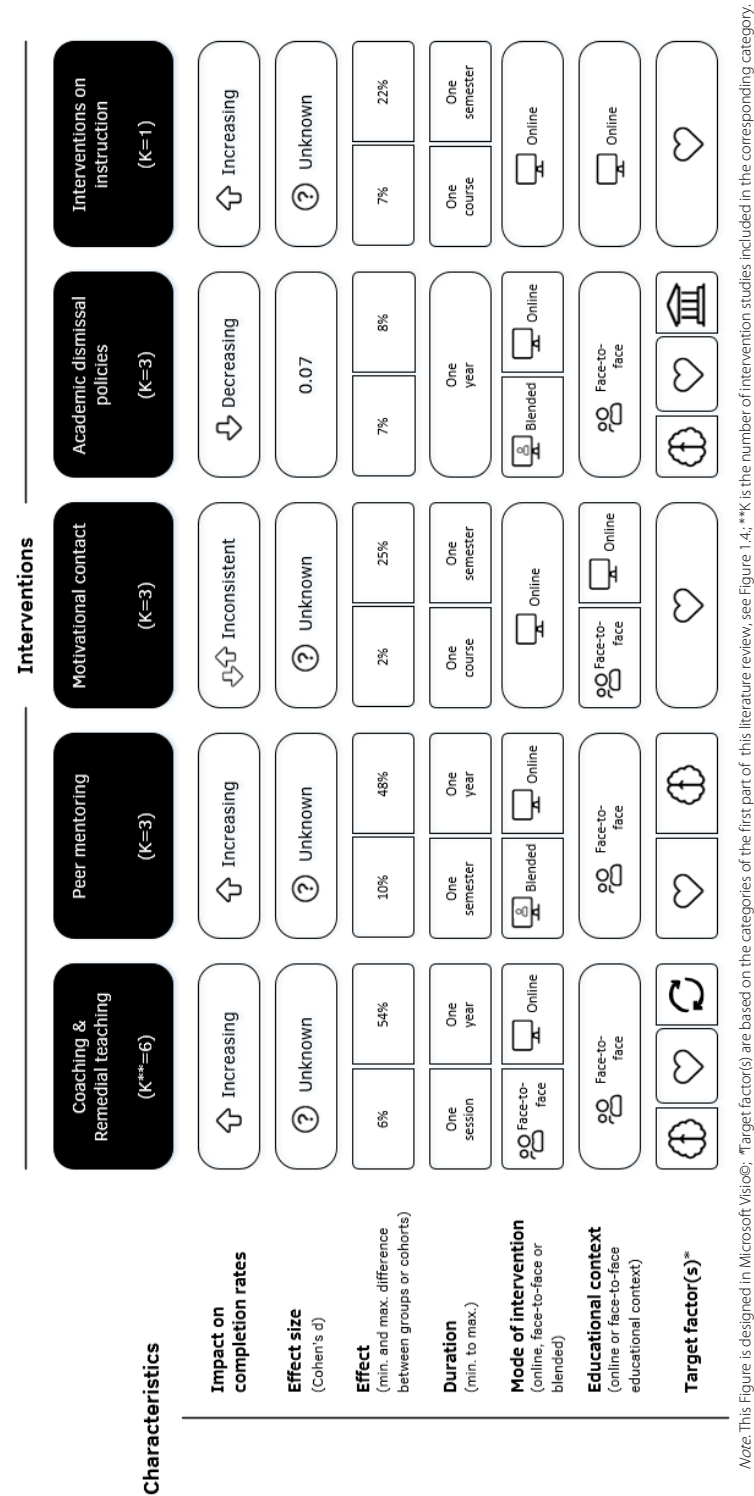
All in all, coaching and remedial teaching interventions aim at increasing completion rates through improvement of mostly dispositional cognitive (e.g. self-regulated learning skills, basic competences such as mathematics) and non-cognitive (e.g. academic motivation, attributions) factors, although one intervention also focused on situational factors (i.e. time commitments outside students' school lives). The interventions in this category included in this literature review are evaluated in the context of traditional higher education mainly. However, Bettinger and Baker (2014) investigated coaching intervention in the context of public, private and proprietary higher education, which also concerns (adult) students combining their study program with a job, comparable to the student population in higher online educational contexts. The majority of coaching and remedial teaching interventions raise complete on rates significantly, although for some of the interventions these results were conditional, for instance, significant in subgroups only. One intervention even decreased completion rates in 2-year community college samples. Coaching and remedial teaching interventions in general have the lowest minimum duration, lasting only one informal session of 30 min, of all interventions included in this review study. Differences between experimental and control groups or cohorts before and after implementation of these interventions of in this category are ranking highest of all interventions in this literature review.

Peer mentoring programs seem to significantly improve completion rates in higher education, but the peer mentoring studies included in this literature review are executed in traditional higher education only, so we cannot draw any conclusions on the impact of these interventions on completion rates in higher online education. The factors focused on in these interventions are merely dispositional non-cognitive (i.e. academic and social integration), although Pagan and Edwards-Wilson (2002) focused on dispositional cognitive factors as well (i.e. study skills). With respect to duration, these interventions with a minimum duration of one semester not as short as the shortest coaching and remedial teaching intervention in this literature review, and last at maximum duration as long as interventions in other categories as well. Effects of these interventions are comparable to the effects of coaching and remedial teaching and motivational contact. One peer mentoring program was stated to be cost effective (Salinitri, 2005), but to draw conclusions on the cost effectiveness of this category, more information is needed, although in comparison to mentoring programs in which students are mentored by university staff or externally hired mentors, peer mentoring might be less expensive.

Motivational contact interventions are mostly implemented and evaluated in the context of higher online education, although in one of the interventions there was a face-to-face condition as well (Huett et al., 2008). These interventions show inconsistent results with respect to increasing completion rates. These interventions are aimed at increasing completion rates through improvement in dispositional non-cognitive factors. According to the researchers of the interventions included in this category of the present literature review, these interventions are cost effective.

Academic dismissal policies, in contrast to the other categories, increase non-completion rates in the first year of higher education. However, by selection after student enrolment these interventions increase graduation rates. However, graduation rates were not the main outcome measure in this review, and therefore makes it hard to compare results of academic dismissal policies to those of the other intervention categories. The interventions in this category included in this literature review have only been investigated in the context of traditional higher education, which means we cannot draw conclusions on whether these interventions do have impact on completion rates in higher online educational contexts as well.

In the category of interventions on instruction, only one study was included in the present literature review, which means we cannot draw conclusions on interventions on instruction in general, as we cannot compare results of different intervention studies in this category. It also makes it difficult to compare the results of this category of interventions to the other categories. The specific study in this category included in the present literature review was carried out in the context of higher online education, but there are, of course, diverse interventions on instruction possible both higher online educational contexts as well as traditional higher education (e.g. education delivery method, problem-based learning).



Note: This Figure is designed in Microsoft Visio®; *Target factor(s) are based on the categories of the first part of this literature review, see Figure 1.4; **K is the number of intervention studies included in the corresponding category.

Figure 1.5. *Synthesis of characteristics of effective and efficient intervention studies*

1.4. Discussion

A vast body of previous research has aimed at explaining the non-completion problem by investigating predictors of non-completion. Interventions aimed at increasing completion rates should be built upon the research on predictors of non-completion (i.e. as this research points out where there is room for improvement), and take into account lessons learned from previously evaluated interventions.

Drawing on 10 review studies, we identified consistent significant effects for the following modifiable predictors of non-completion in higher education: entry qualifications, study- or learning strategies, academic adjustment/adaptation and involvement, goals and intentions, academic self-efficacy, employment, supportive network and faculty-student interaction. Based on effect sizes and consistent results only, the most effective modifiable predictors for non-completion based on these review studies seem to be study-/learning strategies or skills (estimated true correlation of 0.366, see Robbins et al., 2004), academic self-efficacy (estimated true correlation of 0.359, see Robbins et al., 2004), academic goals and intentions (estimated true correlation of 0.340, see Robbins et al., 2004). Estimated true correlations of 0.206–0.230 were reported for academic adjustment/adaptation and involvement (Credé & Niehorster, 2012; Robbins et al., 2004).

Drawing on 16 intervention studies, we found that interventions significantly increasing completion rates in higher education are coaching (i.e. motivational), remedial teaching on basic competences as mathematics and writing skills, and peer mentoring. These interventions focus on dispositional cognitive factors, dispositional non-cognitive factors and situational factors. Inconsistent results were found for the effectiveness of motivational contact on completion rates in higher education. Additionally, results on the effectiveness of interventions on instruction are hard to interpret, as actually, multiple interventions were evaluated at once and there was only one intervention study included in this category. Academic dismissal policies appear to decrease completion rates in the first year, but thereby increase completion rates in the long run. In terms of the possible strategies to raise completion rates, as discussed in the introduction, we could say that academic dismissal policies constitute a third strategy: reducing 'enrolled numbers' by making a definite calculation of enrolled numbers only after the first year. However, such an approach is not without ethical implications, as it might be possible to draw conclusions on the risk of not completing a course or study program earlier in the first year of higher education, or even before student enrolment (e.g. see Fonteyne & Duyck, 2015), when prospective students or newly enrolled students are still in a position to adjust their study decision and/or study behaviour. Providing valid tools (e.g. diagnostic assessments) prior to student enrolment identifying 'weak spots' or barriers affecting chances for completion, with the aim to enable students to better prepare and/or adjust their decision (rather than select), might be a solution in that direction. Especially in distance universities, often characterized by an open access policy combined with high demands in terms of motivation, discipline and self-regulation, 'expectation management' is the predominant aim, rather than selection. Obviously, independent of questions of selection and timing, any diagnostic tools should be developed and

deployed with great care so as to rule out, false negatives, i.e. the possibility of discouraging or dismissing students who in fact stand a real chance of completion.

Interesting is that of the 16 intervention studies included in the present literature review, only four were fully targeted at what appeared to be modifiable consistent predictors of non-completion in the first part of our literature review. Of the 16 included intervention studies, 14 at least partly targeted those modifiable consistent predictors of non-completion. However, the majority of these 14 merely focused on other factors that were not demonstrated as modifiable consistent predictors of non-completion in our review. There is thus only limited explicit alignment between the research on interventions aimed at raising completion rates in higher education and significant relevant predictors of non-completion in higher education as demonstrated in prior research. As stated by Chyung (2001), we also believe that a more systematic approach is needed to resolve the non-completion problem, which means that educational institutions systematically design and carry out interventions. This advocates for a more design-based research approach (Van den Akker et al., 2013), in which interventions are based on results of a needs-assessment first (i.e. take into account where there is room for improvement and what modifiable factors are relevant) and are evaluated in a cyclic manner. As Martens (2018) states, there is a need for educational research in which there is a systematic collaboration between different stakeholders (e.g. practitioners and researchers). To solve complex problems such as the non-completion problem in educational practice, educational research should be conducted in the context in which a solution for the problem is demanded. The interventions carried out in the context of higher online education showed either inconsistent results (i.e. motivational contact interventions) or were hard to interpret, as multiple interventions were investigated simultaneously and in one study only (i.e. interventions on instruction). The interventions that have been found to significantly increase completion rates are carried out in the context of traditional higher education only, even though the mode of the interventions in some cases was (partly) online. These interventions merit further investigation to establish their possible impact in the context of higher online education, and under which conditions.

1.4.1. Limitations of the present review and directions for future research

First, although we gained insight in relevant modifiable predictors of non-completion and to some extent, characteristics of interventions raising completion rates, using these results to develop interventions to raise completion rates in higher education, and higher online education specifically should, be made with some caution. With respect to predictors of non-completion, we cannot simply assume that the same factors apply to the same extent in higher online education as they do in traditional higher education. This means that interventions based on factors relevant for traditional higher education, might be less effective (or not effective at all) in higher online education, for instance due to the higher proportion of adult learners, for whom situational factors (e.g. job, parenting) might play a much bigger role. This is something to take into consideration not only in explaining the non-completion problem, but also in designing interventions for different target groups. Moreover, the way factors should be operationalized possibly

differs between educational contexts. The construct motivation in a group of prospective traditional higher education students, currently leaving high school, is likely to be different in a group of adult learners, picking up a study course or program years (or maybe even decades) after leaving high school, combining it with several other roles, such as a job or parenting.

Second, there are several points that should be taken into account when designing and evaluating new interventions aimed to raise completion rates in higher education. First, in contrast to the vast body of literature on predictors of non-completion in higher education, scientific literature on interventions is rather scarce, in particular on interventions prior to enrolment. In result, a publication bias might have influenced our results, as we only included peer-reviewed articles published in academic journals. Especially for the overview of interventions this might have played a role, as a lot of educational institutions do design and evaluate initiatives to raise completion rates, but do not publish about the outcomes of these initiatives in (peer-reviewed) scientific journals (Simpson, 2010). Instead these initiatives appear more likely to be reported about in popular scientific journals (for an example, see Schlusmans & Winkels, 2017), or in policy documents. In addition, as mentioned in the introduction, in the field of study success there is often a broad range of terms used for a similar concept or definition. This means our search terms might not have been all encompassing in finding relevant peer-reviewed scientific articles on interventions to raise completion rates. In addition, it is noticeable that we did not find any articles in regard to learning analytics, as it has been argued that learning analytics are promising in predictive research and it is an emerging field in interventional research (Rienties et al., 2016). For both parts of the present literature review, it might be that our search terms have not been all encompassing in this respect, as we did not look for learning analytics studies in particular. For the part on predictors of the present review, we focused on prior review studies only. In that sense, it might be that there are no review studies involving learning analytics yet, meeting our inclusion criteria. For the part on interventions in the present review, we searched for interventions purposely designed to increase completion rates in higher education. A great deal of work and research has been done in the field of learning analytics. However, the translation into concrete and theoretically grounded interventions has yet to be made (Jivet et al., 2017).

Finally, the interventions included in the present literature review are all interventions implemented after enrolment. Despite an additional literature search with search terms for pre-enrolment interventions specifically. These kinds of interventions are, for instance, initiatives to make sure students start a study course or program with the appropriate expectations or an appropriate level of knowledge and skills in order to raise the odds for completion. This does not necessarily mean that these kinds of interventions are not implemented by educational institutions. It does, however, indicate that they have received less academic attention than interventions implemented after enrolment.

The focus of future research should be twofold. Future research should focus on how the predictors of non-completion (in interaction with each other) explain the non-completion problem, especially with respect to a higher online educational context. Additionally, (isolated) factors should be targeted in interventions to investigate the effectiveness in raising completion rates in higher education. Moreover, intervention

studies included in the present literature review were limited in their alignment with research on predictors of non-completion. To build a knowledge base on resolving non-completion, more academic attention is required for interventions aimed at raising completion rates in higher education, especially interventions implemented prior to student enrolment. Besides, more systematic research is needed to design, implement and evaluate interventions to raise completion rates incrementally, step by step, based on prior research on predictors of non-completion.

1.5. Conclusion

In conclusion, non-completion in higher education is a multidimensional problem, in which a broad range of (modifiable) factors provide some predictive value, depending on the context in which they are investigated. Although the results are still rather indefinite and results are somewhat limited for the context of higher online education, we gained relevant insight in modifiable predictors of non-completion in higher education. In specific, Entry qualifications, study- or learning strategies, academic adjustment/adaptation and involvement, goals and intentions, academic self-efficacy, employment, supportive network, and faculty-student interaction appear to be relevant modifiable predictors of non-completion in the context of higher education.

In addition, we gained insight in characteristics of interventions increasing completion rates in higher education and the extent to which these interventions are calibrated with research on predictors. Coaching, remedial teaching, and peer mentoring appear to be very promising. However, the alignment with research on relevant predictors of non-completion is limited. Furthermore, interventions require further investigation in the context of higher online education and more academic attention is required for interventions prior to student enrolment.



Chapter 2

▮ Predicting completion: The road to informed study decisions in higher online education

This chapter is based on: Delnoij, L. E. C., Janssen, J. P. W., Dirkx, K. J. H., Gijssels, H. J. M., De Groot, R. H. M., Neroni, J., De Bie, M., & Martens, R. L. (2021). Predicting completion: The road to informed study decisions in higher online education. *Frontiers in Education*, 6, 668922. <https://doi.org/10.3389/feduc.2021.66892>

Non-completion is an ongoing concern in higher education, and even more so in higher *online* education. One way to address this problem is to take initiatives prior to student enrolment, enabling informed decision-making. In line with this, an institution for open higher *online* education seeks to develop a (non-committal) online self-assessment for prospective students. To identify variables (tests) to be included in this self-assessment, the present study aims at validating variables—previously identified as “predictive of completion in higher education and open to intervention before enrolment”—within the current higher *online* education context. Results of correlational analyses indicate that the following modifiable variables are relevant to include: hours planned to study, employment hours, study intention, discipline, discipline confidence, basic mathematical skills, and social support. Based on a sensitivity cut-off of 95% (to minimise false negatives) about 13% of the actual non-completers could be identified correctly. Implications for future development of the self-assessment are discussed.

2.1. Introduction

Higher *online* education is expanding (Seaman et al., 2018). But the openness and flexibility of such online delivery, compared to traditional higher education (e.g. face-to-face education), comes at a price. That is, the number of students not completing a course or program in higher *online* education is impending, despite diverse initiatives taken by educational institutions (Rovai, 2003; Simpson, 2010, 2013; Vossensteyn et al., 2015). It is important for both the student and the educational institution that non-completion is kept to a minimum. For the institution, non-completion amounts to wasted effort (time and money invested), and possibly reputational damage, as completion is often one of the performance criteria presented in catalogues for prospective students' study decisions, and in some countries funding for educational institutions depends on such outcomes as completion rates (Vossensteyn et al., 2015). For the student, non-completion is also an issue in regard to the invested time and money. In addition, (repeatedly) concluding that the chosen study path does not fit one's characteristics (e.g. knowledge, skills and goals) and/or situation (e.g. combining a study with other responsibilities) might have a demotivating effect for future studies. One way to address this problem is to take initiatives (interventions) prior to student enrolment, to help students choosing a study program that optimally suits them, and to ensure that prospective students' expectations with regard to their courses or studies are realistic (Menon, 2004; Oppedisano, 2009; Vossensteyn et al., 2015; Muljana & Luo, 2019). One course of action in the direction of such an improved orientation prior to student enrolment is the development of (non-committal) self-assessments. It has been assumed that non-selective, but adequate and personalised information will help prospective students to make an informed study decision (McGrath et al., 2014). We define these kinds of self-assessments as *"the active participation of students in making judgments about their own characteristics (i.e. knowledge, skills, and expectations), in order to foster reflection on the extent to which these characteristics fit with studying in a specific context"* (definition adapted from Dochy et al., 1999, p. 334). However, interventions aimed at decreasing non-completion implemented prior to student enrolment are not yet strongly flanked by scientific research (Delnoij et al., 2020). To the extent that interventions prior to student enrolment are systematically researched, these studies largely took place in the context of traditional higher education, typically characterized by a target group of students enrolling right after obtaining their high school degree (Fonteyne & Duyck, 2015).

We aim to design and develop a non-committal online self-assessment (i.e. hereafter referred to as "the self-assessment" or SA) to inform prospective students in open higher *online* education about the match between their characteristics (e.g. knowledge, skills, and expectations), and what is conducive to study in higher *online* education. This SA is aimed at identifying prospective students with lower chances for completion and provide feedback on how they can enhance their chances for completion. It will be non-committal, as prospective students will not be obliged to fulfil this assessment, and students will not be selected based on their self-assessment results as the institute operates according to an open access policy. To determine the relevant variables to be included in such a self-assessment (i.e. to ensure evidence-

informed study decision making), we investigated predictors of non-completion in higher education through a review of reviews (Delnoij et al., 2020). As a second step, it is important to verify predictors resulting from that study in the current context for which the SA is being developed.

The aim of the present study, therefore, is to validate the use of previously identified predictors for completion in the context of higher *online* education and to examine which of these variables need to be included in the SA as constituent components. Theoretical considerations underpinning the validation process are elaborated in Theoretical Framework. The (selection of) possible predictors included in this validation study will be discussed in sections 2.2.2. and 2.2.3. The results of this study will be used to develop the SA.

2.2. Theoretical framework

2.2.1. Validation as a process

Validity can be defined as an “overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment” (Messick, 1989, p. 2). In line with this definition above, as well as modern validity theories, we consider validity to refer to the inferences (interpretations and actions based on assessment scores) rather than the instrument itself, and validation as a process requiring ongoing evaluation of evidence, rather than a “once and for all” conclusion (Royal, 2017). In this respect, the validation process described in this paper must be considered as “first steps” of evidence collection concerning the validity of inferences supported by the self-assessment. The Standards for Educational and Psychological Testing developed by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education, mention five sources of evidence (Cizek et al., 2010; Creswell, 2014): evidence based on the test content (e.g. expert opinions), evidence based on response processes (e.g. interviews with test-takers on their experiences with the test), evidence based on the internal structure (e.g. dimensionality and internal consistency), evidence based on relations to other variables (e.g. the predictive value on an expected outcome), and evidence based on the consequences of testing (either intended or not, positive or negative, etc.). The latter type of evidence, according to St-Onge et al. (2017) can be evaluated from both an individual and a societal perspective. In the context of the current self-assessment, the anticipated individual consequence involves the decision to enrol (or not), based on the test scores and feedback. At a societal level, the implicated consequence is a positive impact of the self-assessment on completion rates. Figure 2.1. illustrates the various evidence sources and their relevance at various stages of the (design-based) development of the current self-assessment. Currently, this process is still at the stage of Analysis. Following a previous selection of evidence on predictors ‘in general’ by means of a literature review, this study investigates the relationship of these predictors with completion in the current context. Once this relationship has been established on the predictor level, it will be further investigated at

the model level (i.e. the prediction accuracy of the combined predictors). Prior to investigating the predictive value of predictors measured by means of (sub)scales, it is important to secure the internal structure and consistency of these variables in the present context.

In the next stages of the development and validation process, further evidence will be collected in regard to test content (e.g. user experiences) and response processes. Regarding the latter, a general point of concern is that self-assessments, i.e. self-report measures, may be subject to all kinds of measurement errors, due to, for instance, inaccurate self-perceptions (Dunning et al., 2004) or social desirable answering (Viswesvaran & Ones, 1999; Niessen et al., 2017). Furthermore, as explained above, evidence with respect to the consequences of testing from both an individual (e.g. enrolment decision based on test scores and feedback) as well as a societal perspective (e.g. impact on completion rates) is required in the future stages of the validation process.

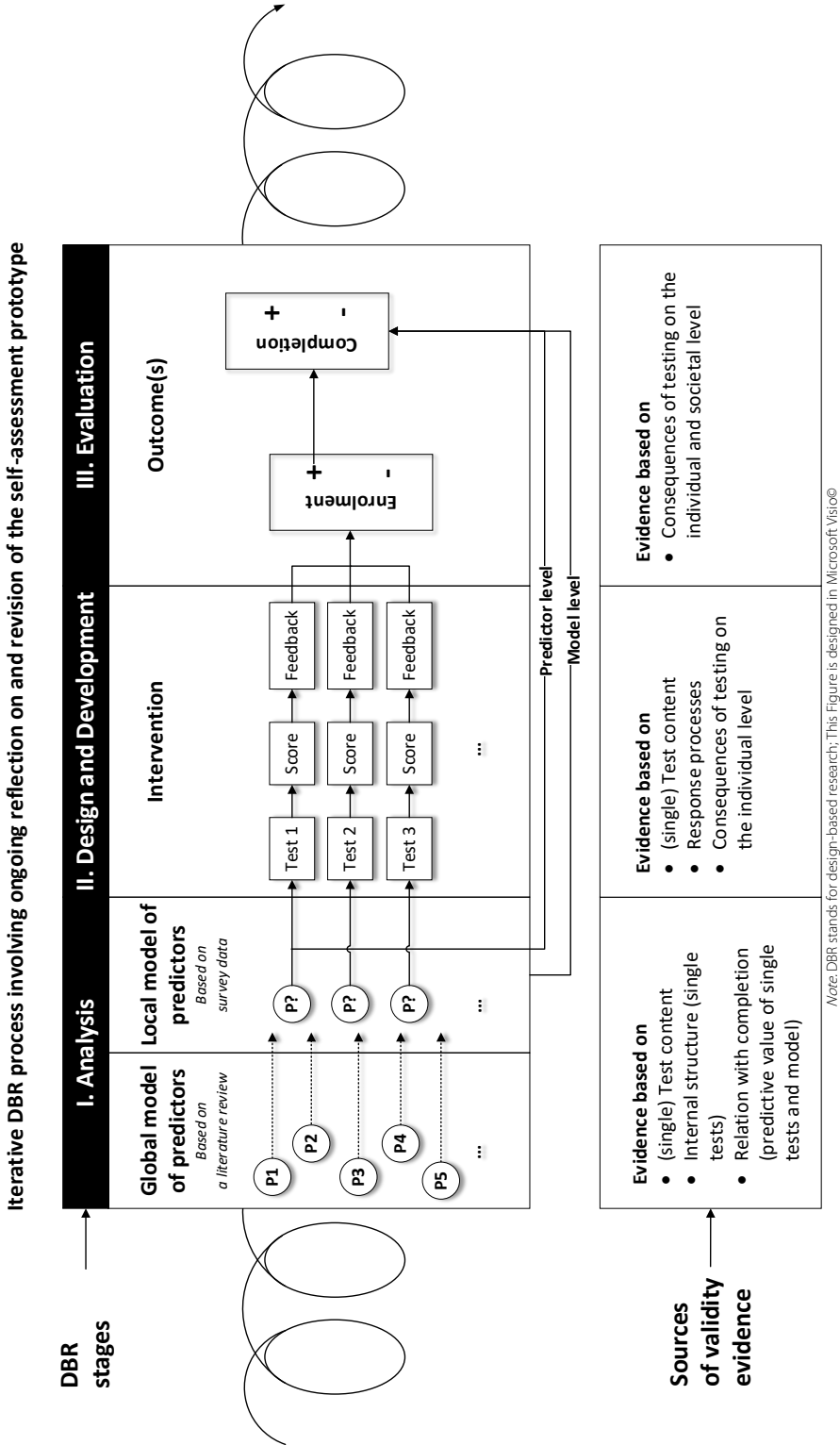


Figure 2.1. Online self-assessment development and validation process

2.2.2. *Predictor selection criteria*

Taking into account that predictors are to be selected as input for a *non-committal/SA prior* to student enrolment, several requirements are formulated to select the possible predictors from prior research. First, predictors need to be identified as variables relevant prior to admission, as it would not make sense to obtain information prior to student enrolment on variables that, in nature, can only play a role after enrolment (e.g. academic adjustment). Prior-to-admission variables identified by Rovai (2003) for instance comprise student characteristics (e.g. academic preparation) and student skills (e.g. time management). Second, the modifiability of variables is a requirement in the present study. That does not mean that non-modifiable variables cannot explain non-completion, or are irrelevant in this context. However, modifiability is required, as the SA needs to give insight into where there is room for improvement and how prospective students can raise their chances for completion. According to our definition, a variable is modifiable if the variable is changeable or can be advised upon. For instance, self-regulation skills (e.g. learning strategies) are trainable (Patterson et al., 2014), and the number of hours a student plans to study can be advised upon (but not be changed directly). Third, as there has been carried out a lot of research on predictors of non-completion in higher education, consistency of prior results is a requirement we take into account. Previous research in this domain is on specific predictors or carried out in specific study programs. Review studies are merely carried out in the context of *traditional*/higher education, and effect sizes are often not reported. This means that conclusions on the predictive value of variables in the context of higher *online* education need to be drawn with caution. We aim to validate predictors from prior research that preferably have been demonstrated consistently (Delnoij et al., 2020). All in all, we aim to include *modifiable* variables, relevant *prior* to student enrolment, which review studies have *consistently* identified as possible predictors of non-completion in higher (*online*) education.

2.2.3. *Selected predictors*

Our previous extensive literature review (Delnoij et al., 2020), yielded potential predictors meeting the selection criteria: academic self-efficacy, employment (hours), basic mathematical skills, study intentions, goal orientation, learning strategies, and social support. Therefore, these predictors are included in the current study. In this previous literature review, motivation, as a possible predictor, surprisingly did not meet the consistency criterion. However, in the context of higher *online* education, it has been argued that a related concept—volition—might actually be more relevant than motivation (Deimann & Bastiaens, 2010). Volition has been defined as “the tendency to maintain focus and effort toward goals despite potential distractions” (Corno, 1994, p. 229). In the case of adults combining a study with a job and family or other responsibilities, distractions or obstacles interfere with the study process. According to Deimann and Bastiaens (2010), motivation might not be enough to overcome these distractions or obstacles. It has been argued that whereas motivation is relevant for initiating activity, volition might be more relevant in accomplishing that certain activity (Deimann &

Bastiaens, 2010). It seems that volition possibly is a relevant variable in relation to our outcome measure of interest. Therefore, we added volition as a potential predictor to our list.

All in all, the current study focuses on the predictors as listed and defined in Table 2.1. The operationalization and measurement of these variables are further elaborated in the methodology section.

Table 2.1. *Definitions of variables*

Variable	Definition	Adapted from
Completion (outcome measure)	The proportion of students enrolling and meeting the requirements for certification, within a specified period of time.	
Academic self-efficacy	The belief in the ability to succeed in an academic environment.	Bandura (1997); Robbins et al., (2004).
Basic mathematical skills	The ability to solve calculations and quantitative reasoning problems.	Fonteyne et al., (2015).
Employment hours	The amount of hours a prospective student spends on paid employment obligations.	
Goal orientation	A reflection of the purpose of achievement behaviour in a particular setting (i.e. academic environment), influencing the way a student approaches academic work.	Harackiewicz et al., (2008).
Hours planned to study	The amount of hours a prospective student plans to spend on studying.	
Learning strategies	Approaches for acquiring, organizing, or transforming information divided in cognitive, metacognitive, and resource management strategies.	Alexander et al., (1998, p. 132); McKeachie et al., (1990).
Social support	Students' perception of whether social networks support them in their academic career financially, emotionally, and practically.	Robbins et al., (2004).
Study intention	The intention to fulfil an educational component (i.e. intention to obtain a master's degree) or not (i.e. orientation or no specific intention).	
Volition	The tendency to maintain focus and effort toward goals despite potential distractions.	Corno (1994, p. 229).

2.2.4. Research questions

The present study aims to gain insight into whether the predictors selected from prior literature are relevant in explaining completion in higher *online* education. Also, we aim to gain insight into the extent to which actual completers and non-completers can be classified correctly by the predictors of non-completion in the context of the Open University of the Netherlands (OUNL), as we want to minimise the risk of falsely

discouraging prospective students. Before we investigate the predictive value of the selected variables and the accuracy of classifying non-completers, it is important to secure the internal structure of predictor operationalization in the current context. Hence, three research questions are subsequently addressed in the current study.

1. To what extent can the internal structure of the instruments used to operationalize the selected predictors be validated in higher *online* education?
2. To what extent is the (relative) predictive value of the selected variables verified by data from a higher *online* educational context?
3. To what extent can prospective students be accurately identified as completers or non-completers by the validated predictors?

2.3. Materials and methods

2.3.1. Context and design

The present study is part of a design-based research process (Van den Akker et al., 2013). As illustrated in Figure 2.1., the results of the analysis phase give input for the design and development of an intervention in a certain context. In the present research, the intervention is the non-committal online SA in the context of the OUNL (i.e. an institute for higher distance education). The OUNL also has to contend with relatively high non-completion rates. For example, approximately 40% of the course participants enrolling in September 2018 did not obtain any study credits within the valid registration period. Education in the OUNL is provisioned mainly online, occasionally combined with face-to-face meetings. Academic courses up to full study programs are provided to obtain a bachelor's or master's degree in the following study directions: law, management sciences, informatics, environmental sciences, cultural sciences, educational sciences, and psychology. The OUNL operates according to an open access policy, which means that for bachelor programs, no prior education is required, and the only requirement is a minimum age of 18 years. Students can choose to study a single course or a combination of courses, up to a full bachelor- or master's program. In general, students have three examination attempts for each course within 14 months after enrolment, after which registration for a course is no longer valid.

The present study can be characterized as a correlational (prediction) design (Creswell, 2014), which means that no conclusions on causality can be drawn from the results. The data is based on two different student surveys, of which the first, most elaborative survey was used to collect data between August 2012 and December 2014 (Neroni et al., 2015). Hereafter, this part of the data is referred to as data collection or dataset 1. As these data did not cover all selected variables, supplementary data collection was executed between September 2017 and February 2019, hereafter referred to as data collection or dataset 2. Data from both data collections were supplemented with data from the student information system on the criterion measure: completion within 14 months after enrolment.

2.3.1. Participants

All (approached) participants were first time enrolling students in the OUNL, as a proxy for the eventual target group of the intervention, prospective students of the OUNL. In Table 2.2., an overview of the sample(s) is provided.

Table 2.2. *Sample information*

	I. Participants approached	II. Respondents ¹	III. Full participating respondents ²	Sex	<i>M</i> _{age} (SD)
				Based on III	
Dataset 1	4945	2562	2043	61.7% Female	43.4 (11.2)
Dataset 2	2996	613	455	52.5% Female	41.2 (11.4)

Note. ¹ Respondents are participants who at least filled in the informed consent and thus, started to fill out the questionnaire; ² Full participating respondents are respondents who filled out the whole questionnaire. For data collection 1, we made a subset of the original dataset including the variables of interest for the present paper. Full participating respondents in data collection 1 are thus respondents who filled out all questions up and until the last question of variables of interest for the present paper, extracted from the whole dataset.

2.3.3. Procedure

The data collection procedure for the two data collections was nearly the same. Newly enrolled students received an email explaining the purpose of the study with an invitation to fill out the online questionnaire. Informed consent was obtained online, preceding the actual questionnaire. Full completion of the questionnaire took approximately 45–60 min in data collection 1 and 30 min for the questionnaire used in data collection 2. Respondents were able to pause and return to the questionnaire if they wished so. Response-enhancing measures included sending out email reminders (both data collections) and follow up phone calls (data collection 1). Besides, in data collection 2, the invitation email was signed by the rector of the educational institution to enhance participation.

2.3.4. Measures

Scale measures. An overview of all independent measures' factors, number of items, and reliabilities (expressed in Cronbach's alpha) based on prior research are given in Table 2.3.

Academic self-efficacy. was measured by the College Academic Self-Efficacy Scale, adjusted by Fonteyne et al. (2017), which we, in turn, adapted to better fit the context of *adult* and *online* learning (i.e. we changed some terms and added three items). The eventual questionnaire consists of 23 items. Respondents were instructed to rate all items on a scale of 1 (completely unable to) to 5 (completely able to). Fonteyne et al. (2017), reported a 2-factor structure with factors identified as effort (e.g. "*Attending class regularly*"), and comprehension (e.g. "*Understanding most ideas you read in texts*").

Table 2.3. *Factor structure per variable and reliability per factor in prior and present research*

Variable (dataset)	Prior research			Present research		
	Factor	K	Reliability (Cronbach's alpha)	Factor	K	Reliability (McDonald's omega)
Academic self-efficacy (d2)	Effort	8	0.760	Confidence in basic study skills	6	0.649
	Comprehension	14	0.790	Discipline confidence	2	0.830
Basic mathematical skills (d2)	Basic mathematical skills	20	0.620	Basic mathematical skills	9	0.722
Goal orientation (d1)	Performance approach	3	0.920-0.960	Performance approach	3	0.880
	Performance avoidance	3	0.820			
	Mastery approach	3	0.880			
	Mastery avoidance	3	0.840-0.890			
	Work avoidance	3	0.900	Work avoidance	3	0.813
Learning strategies (d1)	Rehearsal	4	0.690	Contact with other students	3	0.856
	Elaboration	6	0.750	Discipline	3	0.704
	Organization	4	0.640	Elaboration	3	0.664
	Critical Thinking	5	0.800	Organization	3	0.779
	Metacognitive self-regulation	12	0.790			
	Time & study environment	8	0.760			
	Effort management	4	0.690			
	Peer learning	3	0.760			
	Help seeking	4	0.520			
	Volitional self-efficacy	8	0.790			
Volition (d2)	Consequence control	6	0.800	Consequence control	4	0.802
	Metacognition	9	0.710	Metacognition	4	0.630
	Emotion control	6	0.640			

Note: K is the number of items.

Basic mathematical skills. Basic mathematical skills were measured by a set of 20 items based on work by Fonteyne et al. (2015, 2017). The test consists of open questions, yes/no questions, and multiple-choice questions. One example item is “If $x/y = 0.25$, then $y/x = ?$ ”. There was no time limit and respondents were not allowed to use calculators, although we could not control for that as the test was fulfilled online.

Goal orientation. Goal orientation was measured by the Achievement Goal Questionnaire developed by Elliot and McGregor (2001), supplemented by the Work Avoidance Scale (Harackiewicz et al., 2008). In total 15 items, equally divided in five categories are measured: mastery approach (e.g. “*I want to learn as much as possible from this class*”), mastery avoidance (e.g. “*I am worried that I will not understand everything in this class as thoroughly as I would want to*”), performance-approach (e.g. “*It is important for me to do well compared to others in this class*”), performance-avoidance (e.g. “*I just want to avoid doing poorly in this class*”), and work avoidance (e.g. “*I want to do as little work as possible in this class*”). In the present study, to fit the *adult* and *online* learning context, the word “class” was replaced by “course”. All items are rated on a scale of 1 (totally disagree) to 7 (totally agree).

Learning strategies. Learning strategies were measured by part B of the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1993), adapted to the *adult* and *distance* learning context (e.g. replacing “class” by “course”). The original questionnaire consists of 50 items divided in nine factors: rehearsal (e.g. “*When I study for this course, I practice saying the material to myself over and over*”), elaboration (e.g. “*When I study for this course, I pull together information from different sources, such as lectures, readings, and discussions*”), organization (e.g. “*When I study the readings for this course, I outline the material to help me organize my thoughts*”), critical thinking (e.g. “*I often find myself questioning things I hear or read in this course to decide if I find them convincing*”), metacognitive self-regulation (e.g. “*When reading for this course, I make up questions to help focus my reading*”), time and study environment management (“*I usually study in a place where I can concentrate on my course work*”), effort regulation (e.g. “*I work hard to do well in this course even if I don’t like what we are doing*”), peer learning (e.g. “*When studying for this course, I often set aside time to discuss course material with a group of students from the course*”), and help seeking (e.g. “*I ask the instructor to clarify concepts I don’t understand well*”). Respondents were instructed to consider these items in relation to the way in which they intend to study. All items are rated on a scale of 1 (totally disagree) to 7 (totally agree).

Volition. Volition was measured by an adapted version of the Academic Volitional Strategy Inventory (McCann & Turner, 2004; Deimann & Bastiaens, 2010). All items were introduced as follows: “*If at any point I notice that I’m not working in a targeted manner and with little concentration, then...*”. This questionnaire consists of 32 items, divided in four factors: volitional self-efficacy (e.g. “*I think about my strengths in order not to get blocked by my weaknesses*”), consequence control (e.g. “*I think about the negative effects of not finishing my tasks or projects*”), emotion control (e.g. “*I try to think about joyful things*”), and metacognition (e.g. “*I reflect on my planning and adjust*”).

the associated goals if necessary"). All items are rated on a scale of 1 (completely not applicable to me) to 5 (completely applicable to me).

Single indicator measures. As single indicator measures do not comprise a full questionnaire, they are taken into account only in relation to research questions two and three. An overview of "single indicators" (except covariates) is provided in Appendix E.

Employment hours. These hours were measured as the number of hours in a paid employment contract. Respondents indicating they were not employed, were given value 0.

Hours planned to study. Hours planned to study was measured by one open-ended question: *"How many hours do you expect to study on average per week?"*.

Social support. Social support is divided into financial, emotional, and practical support and measured by newly developed questions, resulting in three dichotomous variables indicating whether respondents receive support (1) or not (0).

Study intention. Study intention was measured by one multiple-choice question in which respondents were asked to indicate their achievement intentions. Their answers were coded 0 if they indicated no specific study intention (i.e. no explicit intention to obtain study credits) and one if they indicated the intention to fulfil a course or program (i.e. in other words, to obtain study credits). Details about these questions and answer options can be found in Appendix E.

Covariates. Covariates taken into account are prior level of education, gender, age, and faculty. These variables were obtained by the student administration office of the educational institution, or inquired by a multiple-choice (e.g. age) or open ended (age) question.

Criterion measure. *Completion* data was distracted from the student information system, with a score of one being assigned to those students completing at least one course within 14 months (after which registration is no longer valid), else a score of 0.

2.3.5. Statistical analyses

Research question 1: To what extent can the internal structure of the instruments used to operationalize the selected predictors be validated in higher online education? To answer the first research question, analyses of descriptives (SPSS Version 24.0, see IBM Corp., 2016), factor analyses, and reliability analyses (Jamovi version 0.9.5.12, see R Core Team, 2018; The Jamovi Project, 2019) were conducted. Prior to factor analyses, items were checked and removed if skewness and kurtosis indicated significant non-normality (Field, 2009; Mayers, 2013; Trochim & Donnelly, 2006). In case no substantial alterations were made to the scale, and sufficiently detailed information was available from prior research, a confirmatory factor analysis (CFA) was carried out. If CFA could not be performed, factor

analyses involved several steps. First, the data was randomly split in half, on which exploratory factor analysis (EFA), followed by an EFA in the CFA framework (E-CFA) (Brown, 2015) on one random half of the dataset. Subsequently, the model resulting from the exploratory analyses was cross-validated by means of CFA using the second half of the data. After that, a CFA on the final model was performed in the whole dataset, of which the results are presented in this paper. We applied relatively strict criteria with the aim to reach an optimal (i.e. most parsimonious) solution, as in the eventual self-assessment, we do not want to burden the respondents unnecessarily. The exact process of and cut-off values applied in factor analyses can be found in Appendix F. For reliability, McDonald's omega was chosen over Cronbach's alpha, as Cronbach's alpha depends on the assumption that each item contributes equally to the factor. McDonald's omega allows items to vary in factor loadings and thereby, fits better to our data. In addition, in using omega there is less risk of overestimation or underestimation of reliability as compared to alpha (Zinbarg et al., 2005; Graham, 2006; Revelle & Zinbarg, 2009).

Research question 2: To what extent is the (relative) predictive value of the selected variables verified by data from a higher online educational context? Analyses regarding this research question started with a check for normality and outliers through descriptive statistics (Trochim & Donnelly, 2006; Field, 2009; Mayers, 2013). Next, various analyses were conducted to gain insight into the relationship between the variables. Pearson correlation coefficients are reported for the relationship between continuous variables. Omega-squared (ω^2) was chosen as the reported effect size for associations between categorical and continuous variables (Analysis of Variance (ANOVA)) because it gives the least biased view on the effect size in analyses in which the assumption(s) of homogeneity of variances and/or normality are not met, which was incidentally the case (Yigit & Mendes, 2018). To decide which categorical variables should be included in the self-assessment, associations between categorical variables (including the outcome measure) are examined using Cramer's V (Cohen, 1988). To decide which continuous variables should be included in the self-assessment, Confidence Interval-Based Estimation of Relevance (CIBER) analyses were conducted in R (Version 3.6.1, see R Core Team, 2018), based on work by Crutzen and Peters (2019). The CIBER analysis was chosen for several reasons. First, it is recommended to base decisions for selecting predictors on confidence intervals for bivariate associations, combined with the variables' distributions and means. Confidence intervals should be used instead of point estimates (e.g. regression coefficients), as confidence intervals give insight in estimation accuracy as well. For instance, a broad confidence interval means that the point estimate is unreliable and can have a substantially different value in a new sample. In the context of selecting variables for the self-assessment, regression coefficients would provide little information on the relevance of specific predictors, because they are conditional upon the other predictors in the model. In regression analyses, it would be hard to distinguish between the contributions of associated predictors in predicting the outcome measure.

Second, CIBER data visualization has two advantages for the selection of predictors in the context of our research:

- It facilitates the comparison of the effects of different variables.
- The relative width of the distribution and variation in estimates is presented, which facilitates a cautious and well-considered decision for variable selection.

Research question 3: To what extent can prospective students be accurately identified as completers or non-completers by the validated predictors? To gain insight into the proportion of explained variance in the outcome measure, the selected variables were included in a multivariate logistic regression together with background variables (i.e. age, sex, faculty, and prior level of education). Given that the SA is constructed to identify those prospective students who have a lower probability for completion, classification accuracy was evaluated in Jamovi (Version 0.9.5.12, see R Core Team, 2018; The Jamovi Project, 2019).

2.4. Results

2.4.1. Internal structure and reliability of scale variables

In Table 2.3., an overview is presented of the results discussed in this section. In Table 2.4., the factor score means, standard deviations, and the minimum and maximum factor scores are presented. For all measurements, the eventual set of items can be found in Appendix E.

Academic self-efficacy. The EFA, E-CFA and CFA procedure resulted in two factors, labelled as *confidence in basic study skills* (6 items, McDonald's omega = 0.649) and *discipline confidence* (2 items, McDonald's omega = 0.830). The correlation between these two factors is 0.178 and significant at the 1% level. This CFA revealed a good fit with SRMR of 0.035, RMSEA of 0.053, TLI of 0.954 and CFI of 0.969 ($\chi^2(19) = 49.2, p < 0.001$).

Basic mathematical skills. Based on prior research (Fonteyne et al., 2015) a CFA was performed in which we examined the fit of a model with one factor including all items. Though the fit of this model was reasonably good (i.e. four out of five fit indices were within cut-off values), there were indications for modifications, and as we aimed for the most parsimonious test, we decided to perform the EFA, E-CFA and CFA procedure. One factor was found, consisting of nine items (McDonald's omega = 0.772). The CFA on the complete data set revealed a good fit with SRMR of 0.026, RMSEA of 0.008, TLI of 0.998, and CFI of 0.998 ($\chi^2(27) = 27.9, p = 0.419$).

Goal orientation. Based on prior research, a CFA was performed (Harackiewicz et al., 2008). Here too, we found indications for modifications, despite a reasonably good fit of the model, so we performed the EFA, E-CFA and CFA procedure. Two factors were found, labelled as *work avoidance goals* (3 items, McDonald's omega = 0.813) and *performance-approach goals* (3 items, McDonald's omega = 0.880). The correlation between these two factors was not statistically significant. The CFA on the complete data

set revealed a good fit with SRMR of 0.010, RMSEA of 0.019, TLI of 0.998, and CFI of 0.999 ($\chi^2(8) = 13.8, p = 0.088$).

Learning strategies. CFA based on prior research showed that the original structure did not fit our sample (i.e. two out of five fit indices within cut-off values). The EFA, E-CFA and CFA procedure resulted in four factors, labelled as *contact with other students* (3 items, McDonald's omega = 0.856), *discipline* (3 items, McDonald's omega = 0.704), *elaboration* (3 items, McDonald's omega = 0.664), and *organization* (3 items, McDonald's omega = 0.779). The model fit of the CFA on the complete data set was good with SRMR of 0.031, MSEA of 0.041, TLI of 0.965, and CFI of 0.975 ($\chi^2(48) = 216, p < 0.001$). All correlations between these factors were significant at the 1% level.

Volition. CFA could not be performed, as the required information was not available. EFA, E-CFA, and CFA resulted in two factors, labelled as consequence control (4 items, McDonald's omega = 0.802) and metacognition (4 items, McDonald's omega = 0.630). The correlation between these two factors was not statistically significant. The model fit of the CFA on the complete dataset was good with SRMR of 0.039, RMSEA of 0.051, TLI of 0.956, and CFI of 0.970 ($\chi^2(19) = 44.2, p < 0.001$).

2.4.2. Predictive analyses

Explorative analysis. Variable means, standard deviations, minimum and maximum scores are presented in Table 2.4., for the two data collections separately. For categorical variables (including the outcome measure), also frequencies are reported. Tables 2.5. and 2.6. provide an overview of relationships between variables, for both data collections separately. Pearson correlations were calculated for relations between continuous variables. None of the correlations exceeds 0.6, and therefore none of the associations is interpreted as high (Evans, 1996). Associations between categorical and continuous variables were examined via Analyses of Variance, of which the effect sizes in ω^2 are reported. Values over 0.14 are considered high (Field, 2009). Table 2.5. indicates several medium-size effects (a ω^2 between 0.06 and 0.14, see Field, 2009). In dataset 1 this is the case for the associations between faculty and age ($\omega^2 = 0.064$), and between sex and organization ($\omega^2 = 0.067$). In dataset 2 the effect sizes on the association between faculty and age, prior level of education and age, prior level of education and basic mathematical skills, and sex and basic mathematical skills are medium-size ($\omega^2 = 0.070, 0.062, 0.074, \text{ and } 0.060$, respectively). Associations between categorical variables (Cramer's V) are presented in Table 2.6. The interpretation of this effect size, ranging from 0 (no association) to 1 (perfect association), is dependent on the degrees of freedom (Cohen, 1988), i.e. the number of possible values of the variable with the least categories. In the present study, no strong associations between categorical variables were found. Medium associations were found between faculty and sex in both datasets (Cramer's $V(1) = 0.378$ and 0.376 for dataset 1 and 2, respectively). In dataset 1 a medium association was found between faculty and prior level of education (Cramer's $V(5) = 0.179$).

Table 2.4. *Descriptive statistics per variable, per dataset*

Dataset 1				
Variable	<i>M</i> (<i>SD</i>)	Min and max	<i>M</i> in % of max	
Performance approach goals	7.22 (3.73)	2.53 – 17.70	40.79	
Work avoidance goals	6.23 (2.68)	2.31 – 16.15	38.58	
Contact with other students	6.85 (3.23)	2.45 – 17.12	40.01	
Discipline	8.19 (2.51)	1.76 – 13.90	58.92	
Elaboration	9.62 (1.84)	1.88 – 13.19	72.93	
Organization	11.09 (2.37)	2.09 – 14.63	75.80	
Paid working hours	27.84 (15.08)	0.00 – 68.00	40.99	
Hours planned to study	12.20 (7.08)	1.00 – 40.00	30.50	
Variable	Count (of respondents)		% of total (respondents)	
Study intention				
a) Specific intention to fulfil an educational component	2269		88.60	
b) No specific intention to fulfil an educational component	201		7.80	
Missing	92		3.60	
Social support: financial				
a) 0 (= no)	1016		35.75	
b) 1 (= yes)	1343		47.25	
Missing	483		17.00	
Social support: emotional				
a) 0 (= no)	71		2.50	
b) 1 (= yes)	2288		80.50	
Missing	483		17.00	
Social support: practical				
a) 0 (= no)	730		25.69	
b) 1 (= yes)	1629		57.31	
Missing	483		17.00	
Faculty / educational program				
(1) Law	569		22.20	
(2) Cultural sciences	261		10.20	
(3) Psychology	903		35.20	
(4) Management sciences	327		12.80	

(5) Informatics	226	8.80	
(6) Educational sciences	113	4.40	
(7) Environmental sciences	102	4.00	
Missing	61	2.40	
Prior level of education			
(1) Elementary school	22	0.90	
(2) Pre vocational education	174	6.80	
(3) Pre higher education	300	11.70	
(4) Vocational education	253	9.90	
(5) Higher education	893	34.90	
university of applied sciences degree			
(6) Higher education	867	33.80	
scientific university degree			
Missing	53	2.10	
Outcome measure			
(a) Completion	1037	40.50	
(b) Non-completion	1525	59.50	
Dataset 2			
Variable	M(SD)	Min and max	Min % of max
Confidence in basic study skills	13.50 (1.27)	8.42 – 16.10	83.85
Confidence for discipline	6.63 (1.13)	3.36 – 8.40	78.92
Basic mathematical skills	2.86 (1.09)	0.00 – 4.24	67.45
Consequence control	9.72 (2.18)	2.84 – 14.18	68.55
Metacognition	7.59 (1.36)	3.51 – 10.89	69.70
Variable	Count (of respondents)	% of total (respondents)	
Faculty / educational program			
(1) Law	101	16.50	
(2) Cultural sciences	89	14.50	
(3) Psychology	143	23.30	
(4) Management sciences	153	25.00	
(5) Informatics	62	10.10	
(6) Educational sciences	27	4.40	
(7) Environmental sciences	28	4.60	
Missing	10	1.60	

Prior level of education			
(1) Elementary school	9		1.50
(2) Pre vocational education	17		2.80
(3) Pre higher education	108		17.60
(4) Vocational education	60		9.80
(5) Higher education	307		50.10
university of applied sciences degree			
(6) Higher education	92		15.00
scientific university degree			
<i>Missing</i>	20		3.30
Outcome measure			
(a) Completion	414		67.50
(b) Non-completion	199		32.50

Note. Descriptives on age and sex can be found in Table 2.2., and are based on full participating respondents.

Table 2.5. Pearson correlation coefficients between continuous variables, w^2 between categorical and continuous variables, per dataset

	1	2	3	4	5	6	7	8	9
Dataset 1									
1 Age		-0.148**	0.038	-0.237**	-0.183**	0.097**	0.157**	0.113**	0.109**
2 Hours planned to study			-0.322**	0.111**	-0.081**	0.022	0.107**	0.005	0.016
3 Paid working hours				-0.070**	0.045*	-0.019	-0.027	0.071**	-0.018
4 Performance approach goals					0.135**	0.123**	-0.127**	0.026	-0.007
5 Work avoidance goals						-0.064**	-0.370**	-0.139**	-0.257**
6 Contact with others							0.100**	0.165**	0.173**
7 Discipline								0.069**	0.205**
8 Elaboration									0.225**
9 Organization									
A Study intention	0.012**	0.010**	0.001	0.000	0.000	0.000	0.003**	0.000	0.000
B Financial support	0.028**	0.002*	0.019**	0.002*	0.000	0.000	0.000	0.002*	0.000
C Emotional support	0.000	0.000	0.000	0.001	0.005*	0.000	0.001	0.000	0.002*
D Practical support	0.007**	0.001	0.000	0.001	0.000	0.011**	0.000	0.000	0.003**
E Faculty	0.064**	0.012**	0.024**	0.006**	0.025**	0.017**	0.009**	0.026**	0.056**
F Prior level of education	0.028**	0.040**	0.020**	0.000	0.001	0.008**	0.003*	0.023**	0.010**
G Sex	0.003*	0.010**	0.022**	0.000	0.007*	0.004**	0.001	0.004**	0.067**
Dataset 2									
1 Age		0.052	-0.095*	-0.002	-0.182**	-0.032			
2 Basic mathematical skills			0.295**	-0.028	-0.133**	-0.027			
3 Confidence in basic study skills				0.139**	-0.032	0.114*			
4 Confidence for discipline					0.056	0.165**			
5 Consequence control						0.090*			
6 Metacognition									
E Faculty	0.070**	0.058**	0.027**	0.008	0.006	-0.001			
F Prior level of education	0.062**	0.074**	0.014*	-0.001	0.013*	-0.007			
G Sex	0.000	0.060**	0.032**	0.004	0.003	0.008*			

Note. * $p < 0.05$, ** $p < 0.01$; Pearson correlations between continuous variables are considered higher from 0.600 or higher (Evans, 1996). The effect size of the associations between categorical and continuous variables, w^2 , is considered low if lower than 0.06, medium between 0.06 and 0.140 and high if higher than 0.140 (Field, 2009). Results from the post-hoc analyses on the associations between categorical and continuous variables are available on request.

Table 2.6. Associations between categorical variables (including the outcome measure) in Cramer's V

	A	B	C	D	E	F	G	H
Dataset 1								
A Study intention								
B Financial support		0.089**	0.045*	0.075**	0.140**	0.074*	0.001	0.066**
C Emotional support			0.122**	0.244**	0.089**	0.081**	0.083**	0.085**
D Practical support				0.215**	0.041	0.069*	0.077**	0.057**
E Faculty					0.074*	0.072*	0.072**	0.062**
F Prior level of education						0.088**	0.378**	0.133**
G Sex							0.101**	0.122**
H Completion								0.003
Dataset 2								
E Faculty								
F Prior level of education						0.179**	0.376**	0.231**
G Sex							0.094	0.248**
H Completion								0.021

Note. * $p < 0.05$, ** $p < 0.01$; The effect size of associations between categorical variables, expressed in Cramer's V. The interpretation of this effect size, ranging from 0 (no association) to 1 (perfect association), is dependent on the degrees of freedom (Cohen, 1988), i.e. the number of possible values of the variable with the least categories. Results from the post-hoc analyses are available on request.

2.4.3. *Selecting determinants of completion*

Dataset 1. The relationship between categorical variables and the outcome measure was examined by means of Chi-square analyses, of which the effect sizes (Cramer's V) are presented in Table 2.6. Study intention, financial support, emotional support, practical support, faculty, and prior level of education show to be significantly associated with completion. However, the effect size of the association between study intention and completion is less than small (i.e. Cramer's $V(1) < 0.10$). Also, the association of financial, emotional, and practical support with completion is very small (i.e. all Cramer's $V(1) < 0.10$). The associations of both faculty and prior level of education with completion, are slightly stronger, but still small (i.e. Cramer's $V(1) = 0.133$ for faculty and 0.122 for prior level of education). CIBER analyses results indicating the association strengths between continuous variables and the outcome measure are presented in a diamond plot (Figures 2.2., 2.3.). The left-hand panel shows the item scores of all participants: in green for completers and in purple for non-completers. The diamonds in the right-hand panel indicate the association strengths (i.e. with 95% confidence intervals). The colour of the diamonds indicates the association direction (i.e. red indicates a negative association, green indicates a positive association, and grey indicates weak associations). The wider the diamond, the wider the confidence interval of the association between a certain variable and completion, meaning that in another sample, a different association between predictor and outcome could be found. Furthermore, some confidence intervals (diamonds) overlap the zero-line (e.g. performance-approach goals and completion), which means that an association of 0 could be a possible outcome as well, in a 95% confidence interval. For dataset 1, the diamond plot (Figure 2.2.) shows that performance-approach goals, work avoidance goals, contact with other students, elaboration, and organization are not strongly associated with completion (i.e. indicated by the grey diamonds, overlapping the zero-line). Hours planned to study and discipline positively associated with completion. Age and employment hours are negatively associated with completion. Of these associations, the association between discipline and completion differs the most from 0.

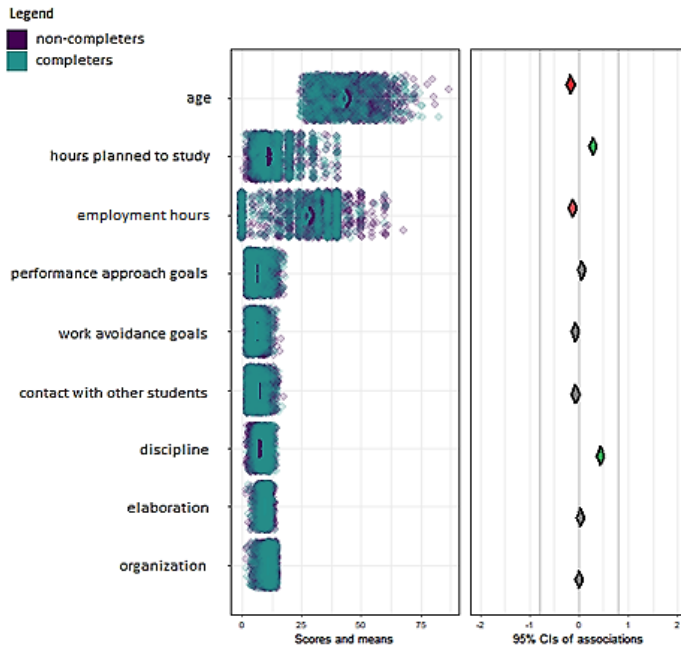


Figure 2.2. Diamond plot of the associations between continuous variables and completion in dataset 1

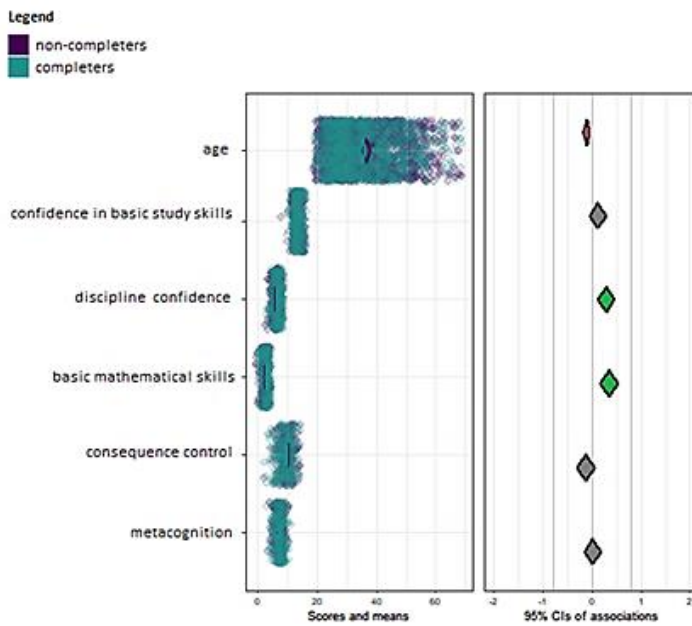


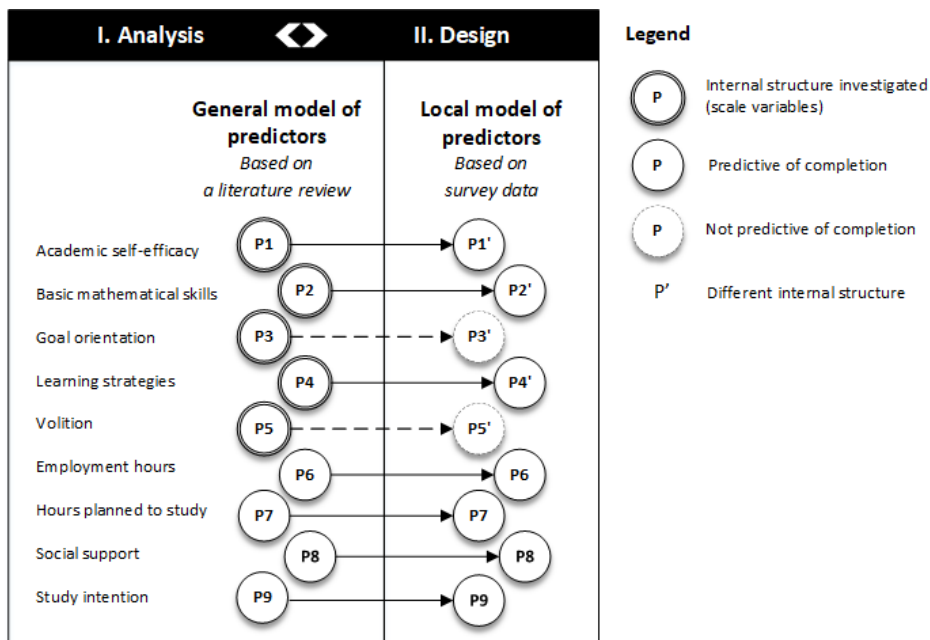
Figure 2.3. Diamond plot of the associations between continuous variables and completion in dataset 2

Based on the Chi-square and CIBER analyses the following modifiable variables are selected for the proposed self-assessment: study intention, hours planned to study, employment hours, discipline, financial support, emotional support, and practical support. To gain insight into the proportion of explained variance, these variables were combined with all background variables (i.e. age, sex, faculty, and prior level of education), in a logistic regression model. Together, they explain 16.3% of the variance (Nagelkerke R^2) in the completion outcome in dataset 1. Excluding the background variables, 8.92% of the variance in completion can be explained by the selected modifiable predictors of completion.

Dataset 2. Faculty and prior level of education show to be significantly associated with completion indicating small to medium effect sizes (Cramer's $V(1) = 0.231$ for faculty and 0.248 for prior level of education). The diamond plot in Figure 2.3. shows that confidence in basic study skills, consequence control, and metacognition are not strongly associated with completion, indicated by the grey diamonds, overlapping with the zero-line in the right-hand panel. Discipline confidence and basic mathematical skills are positively associated with completion, in which the association between basic mathematical skills and completion differs the most from 0. Age is negatively associated with completion, as was the case in dataset 1. Note though, that the confidence interval is not far from including 0. Based on Chi-square and CIBER analyses discipline confidence and basic mathematical skills have been selected as (modifiable) predictors. These variables were added, together with all background variables, in a logistic regression model. Together, they explain 21.7% of the variance in the completion outcome in dataset 2. Excluding the background variables, 7.62% of the variance in completion can be explained by the selected modifiable predictors of completion.

2.4.4. Resulting local model of predictors of completion

Figure 2.1. (see theoretical framework) described the development and validation process of the current self-assessment. As explained, this study focused on collecting validity evidence regarding the internal structure of the selected variables in the current context (research question 1), as well as their relations to completion (research question 2), to establish a local model of predictors. Figure 2.4. zooms in on the analysis stage of Figure 2.1. to clarify the 'filled out' local model of predictors, resulting from the analyses described so far. Predictors requiring an investigation of internal structure evidence are indicated with a double contour in the general model in Figure 2.4. As it happens, the internal structure of all these predictors appeared to differ in the present study, as with an accent mark for these predictors in the local model of predictors. Of the nine variables, originally identified as predictors in the general model (Delnoij et al., 2020), seven are verified (at least partly) as predictors in the current context, and hence, included in the local model of predictors: academic self-efficacy, basic mathematical skills, employment hours, hours planned to study, learning strategies, social support, and study intention. Two variables, goal orientation and volition, are not verified as predictors of completion in the local context (indicated by the dotted arrows toward and the dotted contour of these predictors in the local model).



Note: This Figure is designed in Microsoft Visio®

Figure 2.4. Resulting local model of predictors of completion

2.4.5. Identification and classification

The self-assessment aims at identifying prospective students with lower chances for completion to provide them feedback on where there is room for improvement and how their chances for completion can be enhanced. In light of the open access policy in the current context, it is paramount that the risk discouraging prospective students who, in fact, would have been successful should be kept at a minimum. In other words, we strive to reduce the likelihood of false-negative predictions, at the expense of an increased likelihood of false-positive predictions. For this reason, high sensitivity cut-offs were required. We explored results for two different sensitivity cut-offs: 95 and 99%, which corresponds to a maximum of respectively, 5 and 1% of the prospective students possibly unjustly classified as non-completers. For the sake of comparability with previous studies, background variables were excluded in these analyses. Using the 95% sensitivity cut-off, 13.78 and 12.87% of the actual non-completers were correctly identified as such in dataset 1 and 2 respectively. Using 99% sensitivity, 3.82 and 2.27% of the actual non-completers were correctly identified in dataset one and dataset two, respectively.

2.5. Discussion

The main objective of this study was to select variables to be included in the self-assessment, as a means to enhance informed decision making prior to enrolment. All in all, this study has led to the inclusion of the following modifiable variables in the

self-assessment: hours planned to study, employment hours, study intention, discipline, discipline confidence, basic mathematical skills, financial support, emotional support, and practical support. In line with findings of the literature (Muljana & Luo, 2019; Delnoij et al., 2020) these variables cover characteristics of both the student him/herself (e.g. discipline), and the students' environment (e.g. social support). The present study's findings on employment are in line with previous studies of dropout in online courses (Lee & Choi, 2011) and higher education in general (Riggert et al., 2006). Additionally, discipline seems to be an important topic in the predictors of completion in higher online education. In the present study, discipline appeared a predictor of completion. This association in the context of higher online education has previously been stressed in survey research (Waschull, 2005) as well as qualitative research (Gaytan, 2013). The present study's results for academic self-efficacy were (partly) in line with findings of review studies in the context of higher (online) education (Lee & Choi, 2011; Bowles & Brindle, 2017). However, it is not clear how academic self-efficacy was operationalized in these review studies. Our results showed that, as a factor of self-efficacy, discipline confidence is associated with completion. Furthermore, the association between mathematical skills and completion is in line with findings using the same measurement (albeit in face-to-face education; cf. Fonteyne et al., 2017), as well as studies in the context of online education, using standardized tests (Morris et al., 2005). Finally, our findings on social support are in line with findings in comparable contexts (Asbee & Simpson, 1998; Park & Choi, 2009; Lee & Choi, 2011). However, some results are not in line with previous research or theories on predictors of completion. For instance, goal orientation (i.e. performance-approach and work avoidance goals) did not appear to be related to completion in the present study. A possible explanation lies in the context of the present study, which is characterized by a merely adult student population, combining a study with a job and/or family responsibilities. In this context, specific intentions, rather than the orientation of one's goals (e.g. oriented to outperform others), might be more important for completion. Though the effect was small, this was also suggested by our results, as we did find an association between study intentions and completion. In that regard, research carried out in the context of MOOCs (i.e. another example of open higher *online* education), demonstrated that intention is an important requisite for completion (Henderikx et al., 2017). Besides that, although performance approach goals (i.e. whether or not students' are oriented at outperforming others) might not relate to completion, they appear predictive for other correlates of students' success, such as grade point average (Neroni et al., 2018). Furthermore, in the present study, no association was found between volition (i.e. consequence control and metacognition) and completion. Theories in the field of distance education suggested that volition might be an important predictor of performance and achievement in this context (Corno & Kanfer, 1993; Keller, 2008). However, to our knowledge, there is no empirical evidence for the relationship between volition and completion, although some evidence exists for a relation between volition and academic procrastination in this context (Ucar & Bozkurt, 2019). Overall, we must note that rather strict cut-off values were applied in factor analysis, as we aimed for the most parsimonious tests, which was explained in the method section. This explains differences in the dimensionality of the measures used in the present study and, in turn, might explain differences in our results, compared to previous research.

The variables selected for the self-assessment, together with background variables explain 16.3% (dataset 1) and 21.7% (dataset 2) of the variance in completion. Whether or not these proportions of explained variance are meaningful, is open to debate. According to Allen et al. (2009), this depends on the practical utility of the test scores. In that sense, we consider these proportions of explained variance meaningful, taking into account the results on prediction accuracy of actual non-completers by selected modifiable predictors of the self-assessment irrespective of background variables: 13.78 and 12.87% (dataset 1 and 2, respectively) with a sensitivity of 95%. These are promising results, especially in comparison with similar research in *traditional*/higher education, in which 3.7% of the failing students were identified correctly (Fonteyne et al., 2017). Note that in both cases the results have been achieved in a total sample prediction. Fonteyne et al. (2017) also investigated the classification results in program-specific contexts. Interestingly, 13.4% of the actual failing students were identified correctly using a program-specific prediction. In relation to the open access policy in the current context, in which we want to avoid unduly discouraging prospective students, we might consider being even stricter in setting a sensitivity cut-off. Therefore, we examined the prediction accuracy of actual non-completers also at a sensitivity of 99%, resulting in 3.82 and 2.27% of non-completers that were classified as such in dataset 1 and 2, respectively. These results illustrate a trade-off in which a higher sensitivity results in less false negatives (i.e. maximally 5% at a sensitivity of 95 vs. 1% at a sensitivity of 99%) but at the expense of correct classification of actual non-completers. Note though that the percentage of actual non-completers - that can be classified correctly with a stricter sensitivity (99%) in the present study-is in line with the results obtained at a more liberal sensitivity of 95% in the context of traditional higher education (Fonteyne et al., 2017). In addition, in evaluating the effectiveness of interventions in education, it is not only important to take into account practical utility, but also factors like, for instance, cost-effectiveness and scalability (Kraft, 2020). Considering the latter, the self-assessment seems a promising intervention.

2.5.1. Limitations and directions for future research

Several limitations are noteworthy in regard to the present study, as they point out directions for future research in this field of study. In regard to the practical application of newly constructed scales, the step from “predictor” to “test” (Figure 2.1.) requires an extra step in terms of collecting evidence on (single) test content.

In light of the selected variables for the self-assessment, elaboration of interaction effects was not the focus of the present study. As these relationships might have implications for practice (e.g. gender differences in the relevance of certain variables for completion), a recommendation for future research is to examine these possible interactions, including specific study programs. As shown by Fonteyne et al. (2017), insight in program-specific relationships between factors and completion might result in better prediction accuracy and might have practical implications for feedback to be provided. Furthermore, a considerable part of the variance in completion remains unexplained. In that respect, there might be other modifiable factors associated with

completion, which can possibly be included in the SA to establish a better prediction of completion, and thereby enhance the validity of the SA. For instance, in a review study by Muljana and Luo (2019), it has been shown that technological skills might be a relevant factor, especially in higher *online* education. Finally, in the next steps of the (design-based) development process of the SA, it is recommendable to include additional measures of actual behaviour (e.g. sample tests) next to self-reported behaviour, to enhance the predictive validity and fairness of the self-assessment (Kuncel et al., 2001; Niessen et al., 2016, 2018; Sackett et al., 2016). Actual (study) behaviour in these sample tests is mimicked by a simulation of representative parts of academic programs in a certain context (Niessen et al., 2016). For instance, such a sample might involve studying literature and/or watching video-lectures, followed by a short exam.

2.5.2. *Implications for practice and research*

Currently, there is a high need and demand for online education, because of the covid-19 pandemic. Accessibility to educational programs will widen further, when universities decide to continue offering (partly) online education after the pandemic (Gomez Recio & Colella, 2020). In that regard, (prospective) students need support in making a well-informed study or program choice. To that end, self-assessments prior to student enrolment seem a promising approach (Kubinger et al., 2012; Lee et al., 2013; Fonteyne & Duyck, 2015; Muljana & Luo, 2019; Nolden et al., 2019). Such approaches aim at achieving optimal alignment of students' skills, motivation and cognitive beliefs on the one hand and required skills and attitudes of a particular educational program on the other hand by raising awareness and providing early remediation (Menon, 2004; Hachey et al., 2013; Fonteyne & Duyck, 2015; Nolden et al., 2019). The present study revealed seven predictors of completion in the context of higher online education that are to be included as subtests in such a self-assessment. By these predictors, about 13% of actual non-completers could be correctly identified (with a sensitivity of 95%). It goes without saying that access to higher education constitutes a sensitive ethical issue, especially in the context of (open) online education. Therefore, development and implementation of self-assessments in this context requires thorough and careful validation, not only of the assessment as an instrument but also of the way it is used and whether it affects the decision-making process as intended (Niessen and Meijer, 2017). In the present study, the content, internal structure and predictive aspects of validity were investigated. Though, validation is not a "once and for all" call and these aspects remain under evaluation (i.e. as the population and educational practice change over time), next steps should focus on the other aspects of validity as well. Prospective students' response processes need to be examined to determine whether the self-assessment is used as intended (Beckman et al., 2005; Downing & Haladyna, 2004). Furthermore, to determine the self-assessment's impact, the consequential aspect of validity needs to be evaluated (Beckman et al., 2005; Cook et al., 2014). This involves investigation of, for instance, the impact on prospective students' study choice certainty, enrolment behaviour and study progress after enrolment.



Chapter 3

▮ Designing an online self-assessment for informed study decisions: The user perspective

This chapter is based on: Delnoij L. E. C., Janssen J. P. W., Dirkx K. J. H., Martens R. L. (2020). Designing an Online Self-assessment for Informed Study Decisions: The User Perspective. In C. Alario-Hoyos, M. J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, S. M. Dennerlein (Eds.), *Lecture notes in computer science: Vol. 12315. Addressing Global Challenges and Quality Education*. Springer. https://doi.org/10.1007/978-3-030-57717-9_6

This paper presents the results of a study, carried out as part of the design-based development of an online self-assessment for prospective students in higher online education. The self-assessment consists of a set of tests – predictive of completion – and is meant to improve informed decision making prior to enrolment. The rationale being that better decision making will help to address the ongoing concern of non-completion in higher online education. A prototypical design of the self-assessment was created based on an extensive literature review and correlational research, aimed at investigating validity evidence concerning the predictive value of the tests. The present study focused on investigating validity evidence regarding the content of the self-assessment (including the feedback it provides) from a user perspective. Results from a survey among prospective students ($N=66$) indicated that predictive validity and content validity of the self-assessment are somewhat at odds: three out of the five tests included in the current prototype were considered relevant by prospective students. Moreover, students rated eleven additionally suggested tests – currently not included – as relevant concerning their study decision. Expectations regarding the feedback to be provided in connection with the tests include an explanation of the measurement and advice for further preparation. A comparison of the obtained scores to a reference group (i.e. other test-takers or successful students) is not expected. Implications for further development and evaluation of the self-assessment are discussed.

3.1. Introduction

The number of students not completing a course or study program in higher online education remains problematic, despite a range of initiatives to decrease non-completion rates (Rovai, 2003; Simpson, 2010; 2013; Vossensteyn et al., 2015). It is in the interest of both students and educational institutions to keep non-completion at a minimum (Vossensteyn et al., 2015). One way to address this problem is by taking action prior to student enrolment, ensuring that the study expectations of prospective students are realistic (Oppedisano, 2009; Vossensteyn et al., 2015). Adequate, personalised information has been shown to help prospective students make informed study decisions (Essig & Kelly, 2013; McGrath et al., 2014) and, by extension, reduce non-completion (Kubinger et al., 2012; Wosnitza & Beltman, 2012). A self-assessment (SA) can provide such information (Nolden & Wosnitza, 2016; Nolden et al., 2019).

The current study contributes to the development of such a SA at an open online university. This SA will be available, online, for prospective students and inform them about the match between their characteristics (knowledge, skills, and attitudes) on the one hand, and what appears to be conducive to (read: predictive of) completion in higher online education on the other hand. The aim of the SA is not to select, but to provide feedback for action, so that prospective students can make a well-considered study choice (Essig & Kelly, 2013; Kubinger et al., 2012; McGrath et al., 2014), based on realistic expectations (Oppedisano, 2009). By following up on feedback suggestions (e.g. for remedial materials) they can start better prepared. However, as Broos and colleagues (2019, p. 3) have argued: “advice may contribute to the study success of some students, but for others, it may be more beneficial to stimulate the exploration of other (study) pathways. It may prevent (...) losing an entire year of study when faster reorientation is possible”. Nonetheless, the SA will be offered as an optional, and (in accordance with the open access policy of the institution) nonselective tool to visitors of the institutional website.

A first prototypical design of the SA (i.e. its constituent tests) was created, based on two prior studies: an extensive literature review and subsequent correlational research (Delnoij et al., 2020; 2021). Both studies were carried out to collect evidence concerning the predictive value of constituent tests regarding completion. However, the predictive value is only one of the five sources of validity evidence, as identified in the Standards for Educational and Psychological Testing (Cizek et al., 2010; Creswell, 2014; Royal, 2017). Another important source of validity evidence is the content of the SA (Royal, 2017), which is the main concern of the present investigation.

There are various reasons to investigate content validity, in addition to the predictive value of the constituent tests. The most important one is that, although previous research may have indicated that a certain test (variable) is a relevant predictor of completion, this does not necessarily mean that users perceive it as useful in the context of their study decision. When it is not perceived as useful, it becomes less likely that prospective students complete the test(s) and use the information they can gain from it (King & He, 2006). The previous argument applies not only to each separate test but also to the overarching SA, i.e. whether the SA is perceived as a useful, coherent and balanced set of tests. Second, validity evidence based on the content of a test is not

limited to the content of the actual test but includes the feedback provided in relation to obtained scores. Regarding this feedback, several design questions remain unanswered.

In short, the general research question addressed in this paper is: *'What are user expectations regarding the tests included in a SA prior to enrolment, including the feedback provided on obtained test scores?'* The next sections will provide some theoretical background regarding the SA and the feedback design, before elaborating on the more specific research questions and the methods used.

3.1.1. Self-assessment model

Figure 3.1. provides the domain model (UML class diagram) of the SA (Warmer & Kleppe, 2001). The Figure illustrates that users attain a score on a predictor (i.e. a test, like basic mathematical skills or a single indicator, like the number of hours occupied in employment). A predictor included in the SA represents either a dispositional characteristic (i.e. pertaining to the student, like discipline) or a situational characteristic (i.e. pertaining to student's life circumstances, e.g. social support) (Delnoij et al., 2021). The score a user attains on a test falls within a particular score range (labelled e.g. unfavourable, sufficient or favourable odds for completion). The exact score ranges (their cut-off points) of the current SA depend on parameters, which are set in the predictive model (Delnoij et al., 2021). For this paper, it suffices to understand that feedback is designed in relation to the score ranges, rather than particular scores. With respect to the exact constituent content elements of the feedback (apart from the obvious score, cf. section. 3.1.2.), the current study is designed to fill in the existing gaps as indicated by the empty boxes in the lower right part of Figure 3.1. These gaps will be discussed in more detail in section 3.1.2.

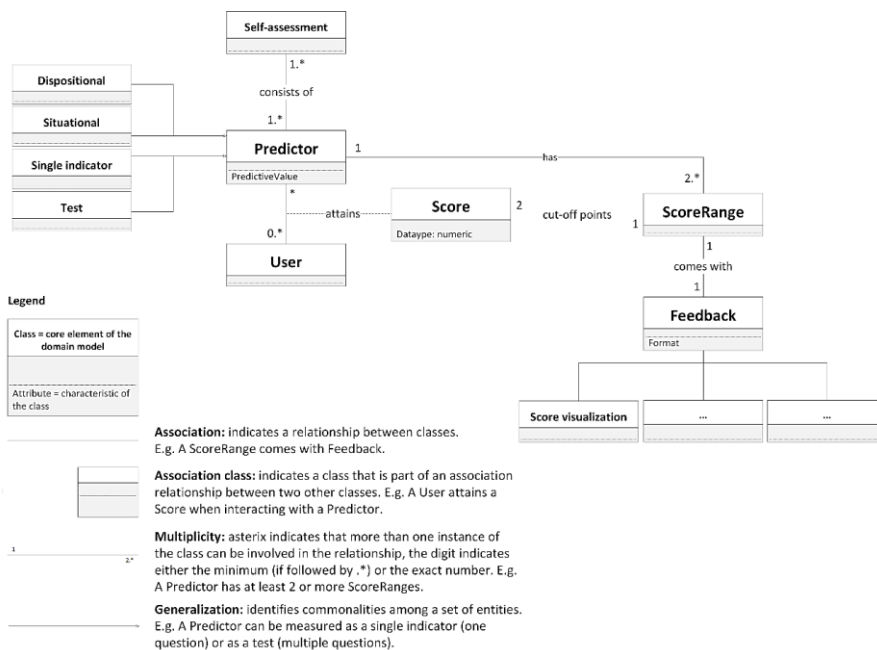


Figure 3.1. *Self-assessment domain model*

Figure 3.2. shows the tests as presented to prospective students in the first prototypical design of the SA. Tests relating to dispositional variables are presented under the headers 'knowledge/skills' and 'attitude'. Situational variables are presented under the header 'profile information'. These headers were chosen, instead of research jargon, to align with the users' frame of reference.

The review study that was carried out to make this first selection of tests was inconclusive regarding a number of predictors and appeared biased towards a face-to-face educational context (Delnoij et al., 2020). This means that, in addition to the tests validated in our previous research (Delnoij et al., 2020; 2021), other tests might be relevant as well. For instance, recent research, not available at the time of the first prototypical design of the self-assessment, has demonstrated that technological skills (e.g. computer skills and information literacy) might be relevant, especially in the context of higher online education (Muljana & Luo, 2019). Furthermore, it has been argued that measures of actual behaviour should be considered next to self-report measures, to enhance the validity of the SA (Niessen et al., 2016; 2018). Actual behaviour might be measured for instance, through a content sample test that involves studying course literature and/or watching video-lectures, followed by a short exam. Such a content sample test has also been shown to predict first-year academic achievement (Niessen et al., 2018). All in all, these are sufficient reasons to collect further validity evidence on the content of the SA so far, and to do so from the perspective of prospective users: if they consider the tests to be useful, they are more likely to complete the SA and use the feedback to help them make an informed decision (King & He, 2006).

Self-Assessment

Die Open Universiteit streeft ernaar dat haar studenten goed voorbereid en goed voorbereid beginnen aan een cursus of opleiding, zodat studenten een goede kans hebben om de studiepunten te een cursus of opleiding te behalen.

Voor deze zelftest zijn een aantal thema's opgenomen, die op basis van een voorafgaande onderzoek zijn geselecteerd. Het is belangrijk dat u weet dat de kennis van deze thema's niet voldoende is om de Open Universiteit, na het inschrijven van de betreffende opleiding, goed te laten weten welke kennis u heeft. Het is belangrijk dat u weet dat de kennis van deze thema's niet voldoende is om de Open Universiteit, na het inschrijven van de betreffende opleiding, goed te laten weten welke kennis u heeft.

Na het invullen van de zelftest wordt u een overzicht gegeven van de resultaten van de zelftest. Het is belangrijk dat u weet dat de kennis van deze thema's niet voldoende is om de Open Universiteit, na het inschrijven van de betreffende opleiding, goed te laten weten welke kennis u heeft.

Meer informatie over de zelftest vindt u op de website van de Open Universiteit.

- Knowledge & skills**
 - Start Basic mathematical skills
- Attitude**
 - Start Discipline
 - Start Study intentions
 - Start Academic self-efficacy
- Profile information**
 - Start Employment
 - Start Social support

Figure 3.2. *First prototype of the self-assessment*

3.1.2. Feedback

Feedback during the transition to new educational contexts has been considered pivotal regarding student motivation, confidence, retention, and success (Nicol, 2009; O'Regan et al., 2016). Feedback on test scores in a study decision process can be designed in various ways (Broos et al., 2018; 2019; Fonteyne & Duyck, 2015; Nolden & Wosnitza, 2016; Nolden et al., 2019). However, with a view on transparency, it is evident that the attained score and an explanation of this score should be part of the feedback. Because the feedback provided on a score is connected to a particular score range (Figure 3.1.), it makes sense to provide and explain the score in this context, as the example presented in Figure 3.3. illustrates.

The attained score is visualized through an arrow in a bar. The bar represents the score ranges. Visualization of feedback data has several benefits as evidenced by research in the field of learning analytics: clearly illustrating a point, personalization, and memorability of feedback information (Sedrakyan et al., 2019). Furthermore, the visualization in a bar representing score ranges is in line with other SAs prior to enrolment (Fonteyne & Duyck, 2015; Nolden et al., 2019).

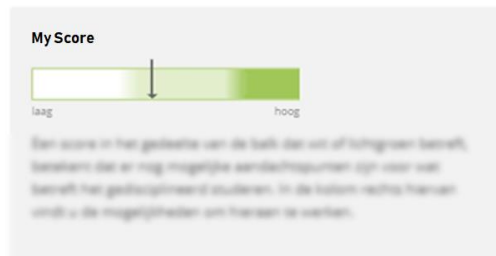
Discipline

Figure 3.3. *In-context visualization of the attained score*

Besides this basic information, additional feedback needs - previously (section 3.1.1.) referred to as gaps - are explored in this study. Current practices illustrate the broad variety of possibilities. For instance, the feedback that is provided in two Flemish self-assessment instruments entailed a comparison of the attained scores to the scores of a reference group consisting of other test-takers (Broos et al., 2018; 2019; Fonteyne & Duyck, 2015) or (successful) first-year students (Broos et al., 2018; 2019). In an online SA used in Germany (Nolden & Wosnitza, 2016; Nolden et al., 2019) the feedback was focused on assisting prospective students in interpreting their scores, independent of comparison to a reference group. What is best, does not become clear when studying the literature. For instance, social comparison theory suggests that in times of uncertainty, individuals evaluate their abilities by comparing themselves to others, to reduce that uncertainty (Festinger, 1954). However, others suggest that information on success or failure in comparison to peers might have an adverse impact on students' motivation and self-esteem (Dijkstra et al., 2008; Nicol & Macfarlane-Dick, 2006).

Another possible feedback component is an indication of the odds for completion, as described by Fonteyne and Duyck (2015). In this case, odds are based on multiple test scores and visualized by a traffic light system. Though students appeared curious about the odds for completion, they also perceived them as quite confronting.

Furthermore, regarding transparency and feedback for action (Jivet et al., 2020), the feedback might contain a description of what was measured (Nolden & Wosnitza, 2016; Nolden et al., 2019) and information for action including tips to improve or a reference to advisory services (Broos et al., 2018; 2019; Nolden & Wosnitza, 2016; Nolden et al., 2019). Regarding feedback for action, Broos and colleagues (2018; 2019) have demonstrated that consultation of a feedback dashboard was related to academic achievement. However, a definite causal relationship with the received feedback (i.e. a change in students' beliefs and study behaviour) could not be established. Broos and colleagues (2019) conclude that dashboard usage may qualify as an early warning signal in itself.

Again, it is paramount that prospective students perceive the feedback as relevant since this will affect their intention to use it, and thereby ultimately, the effectivity of the SA (King & He, 2006). The present study, therefore, investigates prospective students' expectations regarding the feedback provided in the SA.

3.1.3. Research questions

In the present study, we aim to complement the evidence for (predictive) validity of the SA with validity evidence based on the content of the SA, as perceived by prospective users. To that end, we chose to perform a small-scale user study, addressing the following research questions:

1. Which tests do prospective students consider relevant in the study decision process?
2. To what extent do tests considered relevant by prospective students overlap with tests included in the current SA prototype?
3. What are prospective students' expectations regarding the feedback provided in relation to the tests?

3.2. Method

3.2.1. Context

The SA is designed, developed, and evaluated in the context of the Open University of the Netherlands (OUNL), provisioning mainly online education, occasionally combined with face-to-face meetings. Academic courses to full bachelor and master programs are provided in the following domains: law, management sciences, informatics, environmental sciences, cultural sciences, educational sciences, and psychology. The open-access policy of OUNL means that for all courses, except courses at master degree level, the only entry requirement is a minimum age of 18 years.

3.2.2. Design

The present study is part of a design-based research process that typically comprises iterative stages of analysis, design, development, and evaluation (McKenney & Reeves, 2018; Sandoval, 2013). More particularly this study is part of the design stage, reporting on a small-scale user study for further content validation of the SA. This study involves a survey design, examining prospective students' opinions (Creswell, 2014).

3.2.3. Materials

Participants' view on the SA content was investigated via two questions. In the first question, a list of 17 tests, including those already incorporated in the prototypical design, was presented. Tests presented in addition were selected based on a consultation of the literature (e.g. Muljana & Luo, 2019; Niessen et al., 2016; 2018) as well as experts in the field. Respondents were asked to rate the perceived usefulness of each test for their study decision on a 5-point Likert scale (completely useless (1), somewhat useless (2), neither useless, nor useful (3), somewhat useful (4), and completely useful (5)).

In the second question, it was explained that the feedback on each test contains the obtained score and an explanation of this score. Participants were asked to indicate which of the following feedback elements they would expect in addition (multiple

answers possible): *an explanation of what was measured* (Nolden & Wosnitza, 2016; Nolden et al., 2019), *their score compared to the score of successful students* (Broos et al., 2019), *their score compared to the score of other test-takers* (Broos et al., 2018; Fonteyne & Duyck, 2015), *an indication of their odds for completion* (Fonteyne & Duyck, 2015), and *advice on further preparation for (a) course(s) or study program, when relevant* (Broos et al., 2018; 2019; Nolden & Wosnitza, 2016; Nolden et al., 2019).

3.2.4. Participants and procedure

In total 73 prospective students were approached to participate and complete the online survey, resulting in 66 valid responses. Participants constituted a convenience sample (Creswell, 2014) of prospective students who signed up for a 'Meet and Match' event for their study of interest, i.e. law or cultural sciences. We opted for this convenience sample, as it consists of prospective students with a serious interest in following a course or study program at the OUNL (as demonstrated by signing up to the Meet and Match event, for which a fee was charged).

3.2.5. Analysis

Survey data was analysed in Jamovi 1.1.8.0. (R Core Team, 2018; The Jamovi Project, 2019). For the usefulness of the tests (research questions 1 and 2), both the mean (the standard measure of central tendency) and the mode were presented. As the measurement level of the data for the first two research questions was ordinal, we based our conclusions on the mode. A mode of 4 (somewhat useful) or 5 (completely useful) was considered indicative of perceived usefulness. In answering research question 3, frequencies were reported for each answer option (see section 3.2.3.).

3.3. Results

3.3.1. Perceived usefulness of self-assessment tests

The first two research questions were aimed at gaining insight into the perceived usefulness of tests. Table 3.1. provides an overview of prospective students' ratings of the tests. The scores (modes) are ranked from high to low. The tests that are included in the current prototype of the SA are indicated by a checkmark in the first column, to facilitate exploration of the overlap between 'ratings of usefulness' and 'currently included' (second research question).

A content sample test and tests on interests, learning strategies, motivation, academic self-efficacy, career perspectives, information literacy, intelligence, language skills, perseverance, prior knowledge, procrastination (discipline), study goals and intentions, and writing skills are considered useful (Mode 4). Not all currently included tests are considered useful by prospective students.

Two tests (basic mathematical skills and social support) yielded a mode of 3.00, which was below our threshold. On the other hand, academic self-efficacy, study goals and intentions, and procrastination (discipline) were perceived as useful (Mode = 4.00).

Table 3.1. *Tests ranked on mode usefulness as indicated by prospective students*

Test ¹	Mode	<i>M</i> (<i>SD</i>)	Min-max
Content sample test ²	5.00	3.87 (1.14)	1.00-5.00
Interests	5.00	3.88 (1.30)	1.00-5.00
Learning strategies	5.00	4.29 (0.86)	1.00-5.00
Motivation	5.00	3.58 (1.37)	1.00-5.00
✓ Academic self-efficacy	4.00	3.58 (1.24)	1.00-5.00
Career perspectives	4.00	3.67 (1.15)	1.00-5.00
Information literacy	4.00	3.92 (1.04)	1.00-5.00
Intelligence ²	4.00	3.84 (1.02)	1.00-5.00
Language skills	4.00	3.76 (1.10)	1.00-5.00
Perseverance	4.00	3.55 (1.28)	1.00-5.00
Prior knowledge	4.00	3.88 (1.05)	1.00-5.00
✓ Procrastination (discipline) ²	4.00	3.84 (1.02)	1.00-5.00
✓ Study goals and intentions	4.00	3.71 (0.99)	1.00-5.00
Writing skills ²	4.00	4.07 (0.89)	1.00-5.00
✓ Basic mathematical skills	3.00	2.53 (1.23)	1.00-5.00
Computer skills	3.00	2.67 (1.18)	1.00-5.00
✓ Social support	3.00	3.00 (1.24)	1.00-5.00

Note. ¹ Check marks indicate the tests included in the prototypical SA; ² Due to a technical error, only answered by 45 respondents.

3.3.2. Feedback content

The third research question aimed at gaining insight into prospective students' expectations regarding the feedback provided in relation to the SA tests. Table 3.2. presents an overview of the potential feedback elements, ranked by the percentage of students that listed each element (high to low). Next to the obtained score and an explanation of this score (i.e. the minimal feedback), 78.8% of the prospective students expect an explanation of what was measured, and 78.8% of the prospective students expect advice on further preparation for (a) course(s) or study, when relevant. Furthermore, 75.8% of the students expect an indication of the chances of completing a course or study. Finally, a comparison with a reference group is not expected by prospective students, as becomes clear from the relatively low frequencies for both comparisons with scores of other test takers (40.9) and scores of successful students (39.4%).

Table 3.2. *Feedback content elements as expected by prospective students (%)*

	% (N= 66)
Explanation of the test (what was measured)	78.8
Advice on further preparation for (a) course(s) or study, when relevant	78.8
Indication of chances of completing a course or study at the OUNL	75.8
Comparison of obtained score to score of other test-takers	40.9
Comparison of obtained score to score of successful students	39.4

3.4. Discussion

The present study aimed to collect evidence for the content validity of the SA by gaining insight into prospective students' opinions and expectations of a SA prior to enrolment and the feedback it provides.

3.4.1. Self-assessment content

In terms of content validity, further evidence is obtained by the present study for three tests that were already included in the current SA: academic self-efficacy, study goals and intentions, and procrastination (discipline). In line with our previous studies (Delnoij et al., 2020; 2021), these tests appear useful for prospective students as well. Furthermore, the results of the present study show that prospective students find information on specific knowledge (i.e. prior knowledge), skills (i.e. language skills, information literacy, learning strategies, and writing skills), and experience (i.e. a content sample test) useful in the process of their study decision. Although such tests did not appear as relevant predictors of completion in our previous studies (Delnoij et al., 2020; 2021), it might be beneficial to (re)consider and further investigate (e.g. their predictive value in the current context) these as possible tests for the SA. Especially since previous research has also stressed the relevance of, for instance, a content sample test (i.e. providing video lectures on a general academic topic, followed by a short exam) to support students in making well-informed study decisions (Niessen et al., 2016; 2018). Finally, our results show that two tests (i.e. basic mathematical skills and social support) – which proved to be relevant for completion in the online higher education context in our previous studies (Delnoij et al., 2020; 2021) – are not necessarily perceived as useful by prospective students. Part of this result (basic mathematical skills) is likely to be an artefact of the specific sample, i.e. prospective students interested in law or cultural sciences. However, bearing in mind that prospective students need to recognize the usefulness of the tests (Delnoij et al., 2020; 2021; King & He, 2006), this also means due attention should be paid to clarifying the relevance of tests included in the SA to prospective students.

3.4.2. Feedback content

Regarding the content of the feedback, results show that potential users of the SA expect an explanation of what was measured, as well as advice on further preparation for a course or study program at the OUNL, when relevant. Prospective students do not expect a comparison of their score to the score of a reference group (i.e. other test takers or

successful students). Overall, these results are in line with evaluations of feedback in LADs. For instance, Jivet and colleagues (Jivet et al., 2020) have shown that transparency (i.e. explanations of the scales used, and why these are relevant) and support for action (i.e. recommendations on how to change their study behaviour) are important for students to make sense of a LAD aimed at self-regulated learning. Following these results, the feedback in the SA domain model (Figure 3.1.) is complemented with information on what was measured and why, and advice for further preparation for a course or study program in the current context. This information is presented under the headers 'Measurement' and 'Advice', respectively.

'Measurement' contains information on the test and the relevance of this test in relation to studying in online higher education (Nolden & Wosnitza, 2016; Nolden et al., 2019). Yang and Carless (2013) have stated that introducing students to the purpose(s) of the feedback is important for feedback to be effective. 'Advice' provides information on potential future actions that prospective students may take to start better prepared (Broos et al., 2018; 2019; Nolden & Wosnitza, 2016; Nolden et al., 2019). In that regard, feedback literature has suggested that good feedback practices inform students about their active role in generating, processing, and using feedback (Nicol & Macfarlane-Dick, 2006).

Based on the results of the present study we decide not to include a comparison of the attained score to a reference group in the current prototype of the feedback. Furthermore, the odds for completion is not included in the prototypical feedback, even though a majority of prospective students appears to expect this. Calculating an indication of the odds for completion requires predictive models capturing the combined effects of predictors for each program within a specific field (Fonteyne & Duyck, 2015). In the current context, where students do not necessarily commit to a specific study program, but can also decide to enrol in a combination of courses of different study programs, including an indication of the odds for completion appears infeasible. Nevertheless, these results provide input for managing expectancies regarding the self-assessment.

3.4.3. Limitations and future directions

Several limitations are noteworthy in regard to the present study, as they point out directions for future development and evaluation of the self-assessment and the feedback it provides. First, the present study involves a relatively small, convenience sample. Participants were interested in specific study domains (i.e. law or cultural sciences), which is likely to have had an impact on certain results (e.g. perceived usefulness of a basic mathematical test). Thus, it would be valuable to extend the current sample with results of prospective students in other fields. Nevertheless, small-scale user studies can be considered part of the rapid, low-cost and low-risk pilot tests, which are an increasingly important instrument in contemporary research, enabling adjustments and refinements in further iterations of the self-assessment and feedback (Broos et al., 2019).

Second, future development of the self-assessment and its feedback should take into account opinions of other stakeholders, most notably student advisors, as their work

is affected by the SA when prospective students call on their help and advice as a follow-up on attained test results and feedback (Broos et al., 2018).

A third recommendation is to further investigate the extension of the content of the SA, by including measurements of actual behaviour through a content sample test (Niessen et al., 2016; 2018). Interestingly, research has shown that a content sample test is not only predictive of academic achievement but apparently, this experience of the content and level of a study program also has an effect on the predictive value of other tests. For instance, Niessen and colleagues (2017), have demonstrated that scores on other tests (i.e. procrastination and study skills tests), taken after the first course (i.e. an introductory course), more strongly predict academic achievement than scores on the same tests taken prior to enrolment. As the SA is meant to be a generic, rather than a domain specific instrument, we aim to develop a program-independent content sample test (e.g. on academic integrity), in the near future.

Finally, the prototypical feedback merits further investigations of e.g. language and tone (Boscardin et al., 2018), the framing of the score (i.e. focus on what goes well vs. focus on points of improvement) (Jug et al., 2019), possible visualizations (Boscardin et al., 2018; Sedrakyan et al., 2019), and last but not least impact, i.e. consequential validity (Delnoij et al., 2021).



Chapter 4

- ▮ Towards an online self-assessment for informed study decisions –
A mixed-methods validation study

This chapter is based on: Delnoij, L. E. C., Dirkx, K. J. H., Janssen, J. P. W., & Martens, R. L. (2021). *Towards an online self-assessment for informed study decisions – A mixed-methods validation study*. [Manuscript in review]. Faculty of Educational Sciences, Open Universiteit.

Informed study decisions are pivotal for retention in higher online education. Hence, a self-assessment to enable informed decision-making is being developed. Though this requires a thorough validation process, some aspects of validity tend to be underreported. To secure decisions based on the self-assessment, five validity aspects should be evaluated: content, predictive value, internal structure, response processes and consequences. Having established satisfactory results for the first three in previous studies, this study reports on aspects that are less commonly addressed: response processes and consequences. Eight prospective students took the self-assessment in an observed think-aloud mode and were interviewed before and after. Results show response processes to be dependent on the type of subtest. The consequential aspect of validity must be considered in the context of decision-making phases. The demonstrated evidence and possible threats to validity are discussed in light of refining the self-assessment and embedding it in counselling practice.

4.1. Introduction

Adequate, personalised information is pivotal for prospective students to make a well-informed study decision, to stay motivated, and to successfully complete their studies (Nicol, 2009; O'Regan et al., 2016; Tinto, 1999; Vossensteyn et al., 2015). Self-assessments prior to student enrolment can provide such information and are increasingly deployed for informed decision-making (Kubinger, et al., 2012; Nolden & Wosnitza, 2016; Nolden et al., 2019). To determine whether such assessment instruments fulfil their purpose, empirical evaluation is necessary, especially since the use of these instruments can have important consequences for individual decision making and student enrolment. However, empirical evidence is often implicit or completely lacking (Niessen & Meijer, 2017). We argue that such self-assessments should be validated explicitly and as fully as (standardized) summative assessments as well that such validations yield important scientific information that can bring the field a step further. For that purpose, with this study, we show one step in the validation process of such a self-assessment in the context of open online higher education.

4.1.1. *Self-assessment for informed study decisions*

Self-assessments for informed study decisions are advisory and informative instruments conducive to self-examination (Hornke et al., 2013). In general, these instruments aim to elicit reflection on study preparedness by informing prospective students about where they stand in regard to the demands of studying in higher education. One example is the Self-Reflection Tool developed by Nolden et al., (2019). In this instrument, prospective students complete tests and receive feedback on, for instance, self-discipline, learning strategies, and emotional stability. In the feedback, respondents get information about how they scored in comparison to other students. In case, the results indicate issues (e.g. lack of self-discipline), access to remediation is offered by topic-specific recommendations and information about university's support services. In another example, prospective students complete similar tests and receive program-specific feedback focused on their chances of success after enrolment (Broos et al., 2018; 2019; Fonteyne & Duyck, 2015). As self-assessments seems beneficial for retention, we also developed such a self-assessment (Delnoij et al., 2020a; 2020b; 2021). This self-assessment entails three categories of subtests (i.e. knowledge/skills, attitude, and social situation), which have shown to be predictive of obtaining study credits in the context of higher *online* education (Delnoij et al., 2020; 2021). Feedback is provided after each subtest and includes concrete tips and opportunities for remediation, to address possible risks for non-completion early (Delnoij et al., 2020b). Note that our self-assessment is generic; it does not differentiate between or provide an advice for specific study directions. Comparable to the examples given above, the self-assessment not committal and not aimed at selecting students. Rather, the aim is to enable informed decision-making (food for thought), and to encourage prospective students to start well-prepared (feedback for action).

4.1.2. *(The quest for) validity*

These aims pose high demands on assessment validity, i.e. do the test scores, the feedback provided in relation to them, and prospective students' interpretations thereof and following actions all match the proposed use of the assessment?

Hence, to develop an effective self-assessment and feedback (hereafter called 'SA'), it is important to collect and evaluate sources of validity evidence. In the literature, five sources of validity evidence can be distinguished (AERA et al., 2014): content, predictive power, internal structure, response process, and consequences (effects). Investigating these five sources of validity evidence is not a 'once and for all' activity, but one that requires regular attention, as student populations and/or educational practice may evolve over time (Messick, 1989; Royal, 2017). However, a chronological order appears to exist when it comes to collecting evidences from these sources: investigating response processes and consequences makes sense only after the content, internal structure, and predictive power have been more or less secured.

So far, applied validation studies tend to mainly focus on the first three (Cook et al., 2014), also in the specific context of study decision making instruments. More specific, for self-assessments prior to student enrolment, the determination of which tests to include (content aspect of validity), their internal structure, and predictive value (e.g. for retention after enrolment) are often theory- and data-driven (e.g. see Nolden et al., 2019). However, scientific attention is lacking for how prospective students actually proceed through such instruments (*response processes*) and how these instruments affect their study decision (*consequences*). To create a complete picture of the self-assessments' effectiveness, these validity aspects cannot be ignored (AERA et al., 2014; Cook et al., 2014).

Having established satisfactory results regarding content, internal structure and predictive aspects of validity in previous studies (Delnoij et al., 2020a; 2020b; 2021), the present study aims to investigate *response processes* and *consequences* of a self-assessment for informed study decisions.

4.1.3. *Process and consequential aspects of validity*

The process aspect of validity. This aspect comprises theoretical and empirical analyses evaluating how well test takers' actions (responses) align with the intended construct (Cook et al., 2014). The focus is on users' response processes, including the *actions*, *thought processes*, and *strategies* of individual respondents while taking the assessment (Beckman, et al., 2005). *Actions* provide insight into whether prospective students use the SA as intended. In the present study, we focus on the selection and order of subtests taken and the extent to which feedback information is consulted. Respondents' actions are often studied through observation (Cook et al., 2014; Goodwin & Leech, 2003). Additionally, by asking respondents to think-aloud, their *thought processes* (i.e. considerations for providing certain answers) and *reactions* (on a specific test or its items) can be investigated by interviews or asking respondents to think-aloud while they are taking the self-assessment (Cohen, 2006; Cook et al., 2014; Goodwin &

Leech, 2003; Kutlu & Yavuz, 2019). In (concurrent) thinking aloud, participants verbalize their thoughts as they complete a task (Van den Haak et al., 2003). This research method has proved a valid source of data about participants' thinking (Charters, 2003). For securing trustworthiness, follow-up interview questions are proposed, to capture as many of respondents' experiences as possible and to validate researchers' interpretations of participants think-aloud verbalizations (Charters, 2003; Padilla & Benítez, 2014).

Using these methods, valid *strategies* to complete subtests can be estimated (Cohen, 2006; Kutlu & Yavuz, 2019; Padilla & Benítez, 2014). This is important as the validity of strategies depends the content and format of a test (Cohen, 2006). For cognitive tests (i.e. testing knowledge or skills, answers are right or wrong), for example, strategies such as cheating and guessing are clearly flawed (Cook et al., 2014). On the other hand, a common valid test taking strategy is to go back to a specific question or item for clarification (rereading or paraphrasing)(Cohen, 2006). Test-taking strategies may also be flawed by specific measurement techniques. Non-cognitive tests (i.e. measuring attitude or affect) involve test-takers to classify themselves in which self-knowledge and experience is called upon. Such self-report measures, in general, are more prone to socially desirable answers, especially in high-stakes contexts (Cook et al., 2014; Niessen et al., 2017). The relative 'low-risk', non-committal nature of the SA can be expected to reduce socially desirable answers. Nevertheless, investigating variations in response processes may reveal relevant evidence for the process aspect of validity and threats in the sense of variance that is irrelevant to the constructs being measured or the purpose of the SA (Downing & Haladyna, 2004). Thus, results gained from studying prospective students' response processes may reveal relevant implications for development and improvement of the SA.

Consequential aspect of validity. A second focus of this study is the consequential aspect of the SA's validity. Though added later as a distinct source of validity evidence, the literature shows that the consequential aspect of validity is solidly embodied in the current Standards (AERA et al., 2014; Downing, 2003). The consequential aspect of validity pertains to anticipated and unanticipated consequences – both positive and negative – of measurement (Cook et al., 2014; Downing, 2003; Goodwin & Leech, 2003), which can support or challenge the validity of score interpretations and actions based upon them (Beckman et al., 2005). Consequence evidence can be evaluated both from an individual and societal perspective (St-Onge et al., 2017). In the context of the current SA, anticipated individual consequences range from interpretations of the scores and feedback to the decision on whether or not to enrol. The extent to which consequences are valid requires interpretation of the context in which the consequences occur. Increased levels of study choice certainty, for example, are a valid consequence if one scores well on the SA. In this particular context, feeling affirmed in an already certain choice can also be considered valid. A valid consequence to a poor score would be (the intention) to take remedial measures as a follow up on the feedback or even to postpone or reconsider the study decision. Though of course, in the context of open education, we want to be particularly careful not to unnecessarily discourage prospective students. At a societal level, the anticipated consequence is a positive impact of the SA on completion

rates. The latter, impact on completion rates, requires ‘mainstream’ deployment of the SA. Prior to the decision for a ‘full release’ of the SA, (i.e. making it available and evaluate it on a large scale), investigating individual consequences will help to shed light on the question whether the anticipated effects of the SA such as taking remedial measures, postponing and/or reconsidering enrolment, and study choice certainty are evoked as intended.

In the present study, the focus is on the consequences of the SA on the individual level. This means we investigate how prospective students respond on obtained scores and feedback, the extent to which they intend to follow up on the feedback they receive, as well as possible impact on their study choice and certainty thereof.

4.1.4. Research questions

The transition and access to higher (online) education requires the best possible support for students in making a study decision. Therefore, self-assessments deployed for that purpose should be thoroughly validated. With this study, we aim to contribute to a standard for such validation processes by zooming in on two aspects of validity that have not received much attention in validation studies so far, but are important in determining the effectivity of such self-assessments (Cook et al., 2014; AERA et al., 2014): response processes and consequences of testing. The resulting evidence and threats to validity provide insight for the (re)design of a self-assessment for informed study decisions. In other words, we aim to answer the following central research question:

What evidence and threats to process and consequential aspects of validity do we find for the self-assessment and what implications does this have for its design?

To answer the central research question, several sub questions are formulated. Questions establishing a baseline/context:

- **RQ1.** *What are prospective students’ expectations regarding the impact of the SA?*
- **RQ2.** *What are prospective students’ obtained scores on the subtests of the SA?*

Questions regarding the response process, i.e. how prospective students proceed through the SA:

- **RQ3.** *Which tests are selected, in what order and which feedback is consulted while taking the SA and why?*
- **RQ4.** *What reactions are elicited while taking the SA?*

Questions regarding consequences: interpretations, intentions, decisions:

- **RQ5.** *How do prospective students respond to obtained scores and the feedback they receive?*
- **RQ6.** *To what extent do prospective students plan to follow up feedback provided, and what reasons do they have for this?*
- **RQ7.** *How does the SA affect prospective students' study choice and certainty thereof?*

4.2. Method

4.2.1. Context

The SA is designed and developed for prospective students of the Open University of the Netherlands (OUNL), which provides academic courses as well as full bachelor and master programs, mainly online, occasionally combined with face-to-face meetings. The open access policy of the OUNL means that the only entry requirement is a minimum age of 18 years (though naturally, additional entry requirements may be formulated for more advanced courses).

4.2.2. Design

The present study represents a particular step in the design-based research approach, typically comprising iterative stages of analyses, design, development, and evaluation (Van den Akker et al., 2013). More particularly, this study evaluates evidence for response process and consequences through a convergent mixed-methods design (Creswell, 2014) involving observation, think-aloud and semi-structured interviews.

Quantitative data were collected through the subtests, observation and the semi-structured interviews. These data include the obtained subtest scores (RQ2), the number and order in which subtests were taken, consultation of feedback (RQ3), and study choice certainty expressed on a scale of 0 (certain not to enrol) to 10 (certain about enrolling)(RQ7).

Qualitative data were collected through think-aloud as well as semi-structured interviews. These data involve prospective students' expectations of SA's impact (RQ1), their reactions on the subtests (RQ4), their response to obtained scores and feedback (RQ5), and their reflections regarding consequences of the SA (RQ6 and 7).

4.2.3. Materials

In this section, we describe the SA (prototype), observation and think-aloud protocol as well as the semi-structured interview protocol.

Self-assessment prototype. The SA prototype, illustrated in Figure 4.1., consists of four constituent tests, completion of which results in a score and related feedback per

subtest. The subtests measure numerical skills², discipline, social support, and hours planned to study (Delnoij et al., 2021). The numerical skills subtest involves nine items in either multiple choice or open-ended formats. One example item is '*Which of the following options is less than 1?*' with five answer options in which respondents have to add two fractions. The discipline subtest consists of three items on a 7-point scale ranging from totally disagree to totally agree. For instance, '*I find it hard to stick to a study schedule.*' Social support entails one item asking prospective students to indicate for three sources of social support (financial, emotional, practical) whether they receive this from their environment (i.e. partner, family, friends, co-workers, and/or employer). Examples for the three support sources are given and respondents can select multiple answers or a 'none of the above'-option. Hours planned to study is measured by a multiple-choice question with categorical answer options such as 0-5 or 6-10 hours per week.

The feedback design is based on related work in other contexts (Broos et al., 2018; 2019; Fonteyne & Duyck, 2015; Jivet et al., 2020; Nolden et al., 2019) and further informed by the results of an initial user study (Delnoij et al., 2020b). The feedback consists of three components: information on the obtained score, information on the test (what was measured and why), and an advice for further preparation (e.g. general tips, services and contact information of study advisors and opportunities for remediating tutorials at the OUNL). Information on the obtained score is communicated by means of a visualization in which the obtained score, indicated by an arrow, is projected on a bar representing the possible range of scores (scale of 0 – 100%). The colour in the bar fades from white ('high-risk' area) via light green ('medium-risk' area) to dark green ('low-risk' area) indicating increased odds of obtaining study credits. After completing a test, the arrow in the bar is presented on the overall self-assessment dashboard, additional feedback information can be consulted by clicking the result button that appears alongside (see Figure 4.1., C-E).

² In previous chapters, we refer to this subtest as basic mathematical skills.

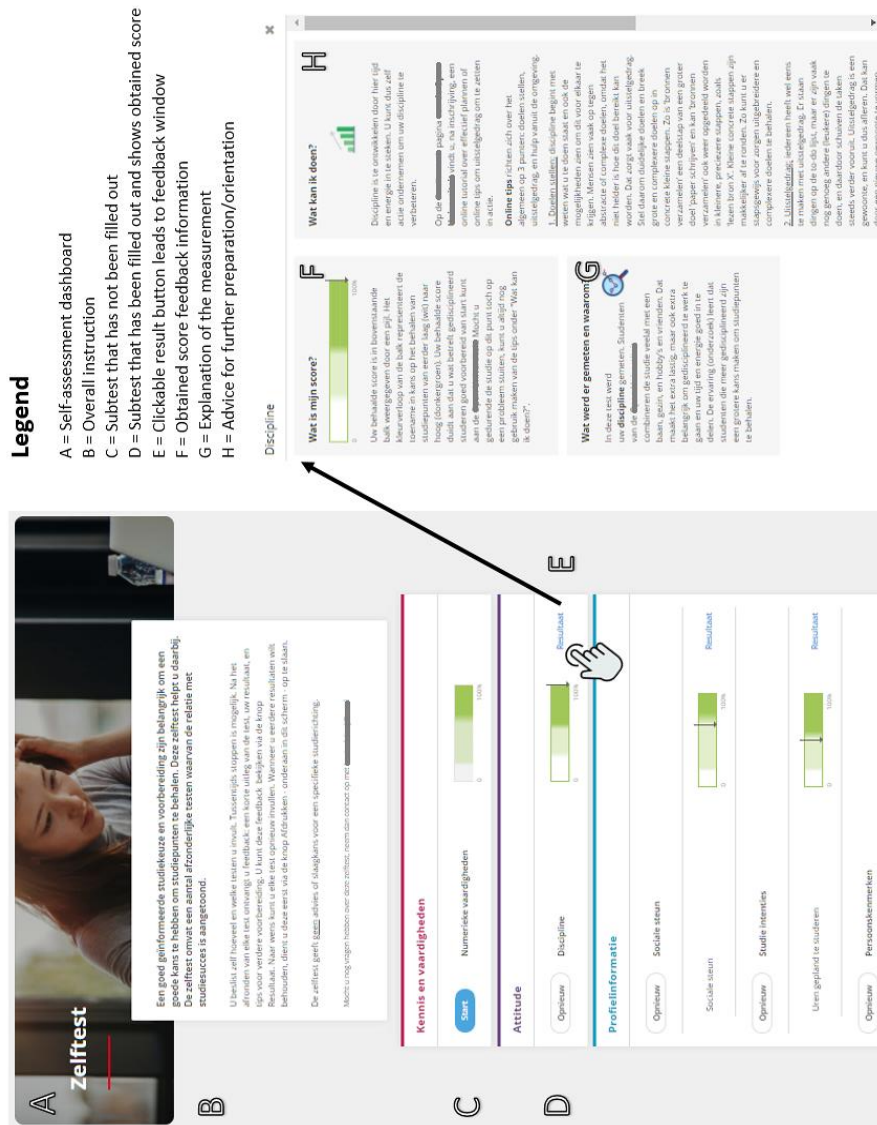


Figure 4.1. *Prototypical self-assessment*

Observation & think-aloud protocol. To observe participants while taking the SA they were asked to share their screen, so that the following actions, related to the process aspect of validity, could be captured: number and order of subtests taken, feedback consultation (i.e. do prospective students consult the feedback or not and, if so, how quickly do they seem to go through it?). A think-aloud protocol was carried out to capture participants' test-taking strategies and reactions while taking the subtests (process aspect of validity) and gain insight into how they respond to their obtained scores and feedback (consequential aspect of validity). We based our think-aloud protocol on previous (related) work (e.g. Charters, 2003; Padilla & Benítez, 2014). In the present study, participants were instructed to express aloud anything coming to mind while taking the SA (e.g. considerations regarding the order in which they filled out the tests, spontaneous feelings and reactions evoked by the test items) and while consulting the obtained score and the feedback provided alongside. Furthermore, it was stressed to participants that it was the SA that was being tested in the present study, not them. Before the actual think-aloud procedure was carried out, it was briefly exercised to allow participants to become familiar with it. The protocol further contained the instruction that in case participants remained quiet for 5 seconds or longer, the researcher should kindly remind them to think-aloud, by asking *'What are you thinking right now?'* The think-aloud procedure stopped when participants indicated that they had finished taking the subtests of their choice. Subsequently, questions were asked to validate the researcher's interpretation of the think-aloud utterances as a source of triangulation (Charters, 2003). After that, the researcher moved on to the interview questions on participants' experiences with the SA as described in the next section.

Semi-structured interview protocol. The interview protocol consisted of instructions for the interviewer (i.e. steps to take prior to the interview), instructions for the participant (e.g. there are no right or wrong answers, try to be as complete and honest as possible in answering the questions), and a list of pre-defined questions on which follow-up questions were asked if necessary. Pre-defined questions were formulated with a focus on both participants' expectancies prior to taking the SA, (e.g. *If so, to what extent do you expect an impact of the SA on your study choice?*) and their thoughts and reflections after taking the SA (e.g. *If any, which follow up actions will you be taking, based on the SA?*). Prospective students' certainty of their study decision was measured on a scale of 0 (certain not to enrol) to 10 (certain to enrol) both prior to and after taking the SA.

4.2.4. Participants

Eight prospective students participated in this study (6 Female, $M_{age} = 36.25$). One participant was interested in following a course, the other seven in following a full study program. Five participants were interested in the domain of law, two in management sciences and one in psychology. All, but one participants already possessed a degree in higher education (university of applied sciences).

4.2.5. Procedure

Sampling procedure. Sampling took place in June and July of 2020. Prospective students who indicated their interest for a course or study program at the OUNL (e.g. by calling the service and information department for information on a certain course) were informed about the study and invited to leave their e-mail address if interested in participating. They received the information letter and link to the online consent form via e-mail. After signing the consent form, an appointment was made.

Research procedure. The sampling procedure was carried out after obtaining ethical approval of the study. In the meantime, a pilot session was conducted to test the research procedure and the latest prototype of the SA. When it comes down to trustworthiness of qualitative research, pilot tests contribute to enhancing credibility and confirmability (Guba, 1981; Krefting, 1991; Shenton, 2004). Based on this pilot session, no adjustments were made for the research protocol. The textual feedback provided with some of the subtests was adapted in order to make it more concise, without loss of content.

The research took place in Blackboard Collaborate®, an online virtual conferencing tool providing functionalities for video calling (i.e. sharing camera and microphone) and virtual lectures (i.e. screen sharing, sharing content). In this session, participants first received explanations on the content and duration of the session. Any additional questions were answered after which the researcher inquired participants' expectations of the self-assessment. Next, the think-aloud procedure was practiced in a mock test very similar to those in the actual SA. Subsequently, participants were instructed login into the online SA environment, upon which the actual think-aloud procedure began. Participants were instructed to notify the researcher once they had taken the tests they wanted to take and read all the information they wanted to read. Afterwards, the follow-up interview took place. Finally, the researcher answered remaining questions and thanked participants for taking part in the study. Participants received a portable document format (PDF) of their obtained SA scores and feedback. All sessions (including the pilot) were recorded (of which participants were informed in the information letter and again during the session).

4.2.6. Analysis

The mixed-methods design of this study involved collection of various data, both quantitative and qualitative. The expected impact of the SA (RQ1), obtained subtest scores (RQ2), total number of subtests taken and feedback consulted (RQ3), intended follow-up actions (RQ6), and study choice certainty (RQ7) are summarized in descriptives. Participants' reactions while taking the SA (RQ4), responses to obtained scores and feedback (RQ5), and further reflections (RQ6 and 7) are analysed using qualitative content analysis.

Qualitative content analysis. As a starting point of the qualitative data analysis, audio recordings were transcribed verbatim. All transcripts were first read in depth to

allow familiarization with the data. Next, an iterative coding process took place. Two researchers coded one part of the data separately first. For securing credibility and confirmability (Guba, 1981; Shenton, 2004), they discussed their coding results together and with a third researcher. Initial categories of codes and themes of categories emerged from this discussion. Based on that, the principal investigator coded the rest of the data. Ambiguities were solved in consultation with the other two researchers. The coding process was carried out in accordance to the steps of qualitative content analysis as described by Erlingsson and Brysiewicz (2017). The first step in that process was to split up the data in (condensed) meaning units: a short text fragment, in which the core meaning is retained. These condensed meaning units were coded. A code is a label that most accurately describes what a condensed meaning unit is about, usually in 1 or 2 words. For example, "It has been a long time since I have had to keep track of such a schedule, so I don't know" was coded as "Lack of recent experience" and "I don't fully trust my own answers" was coded as "(Possibly) flawed answering". After that, codes were grouped into categories, e.g. a group of codes that are related to each other through content or context and is usually factual and short. For instance, the codes "Lack of recent experience" and "(Possibly) flawed answering" were grouped together as "Process threat". Subsequently, we inspected categories to elicit the main themes. These themes express an underlying meaning of two or more categories, and are descriptive in name. As an example, "Process threat" and "Process evidence" were grouped together as "Process aspect of validity".

4.3. Results

4.3.1. *Expectations of SA impact (RQ1)*

Table 4.1. provides a summary of whether or not an impact of the SA on study choice was expected. Four participants did not expect the SA to have much impact on their study decision e.g. because they already had gone through an extensive orientation process, expressed as "I would say the assessment will not have much influence on my decision, as I already did a lot of research" (participant P³). Nevertheless, it can help improve their understanding of what studying in the specific educational context will entail. Participant L mentioned this as following: "I will definitely continue the study decision I already made, but then at least I will have a better picture of the time and effort it would cost me."

Four participants expected the SA to have an impact on their study decision in the sense that they are seeking affirmation on whether or not they are making the 'right' decision. Participant J said, "That I get a kind of confirmation whether or not my decision is a good idea" and participant E stated "Either a confirmation of what you already have in mind or of your insecurities and, therefore, a confirmation to look further and choose something else".

³ To secure anonymity, participants were given a random identifier.

Table 4.1. *Overview of expected impact, test taking behaviour, obtained scores, feedback consultation and study choice certainty*

	Participant							
	J	P	L	Y	I	K	E	Z
Impact on study choice expected	yes	no	no	no	no	yes	yes	yes
Test taking order¹ obtained score² and feedback consultation³ per subtest								
Numerical skills	1 ✓	4 ✓	1 ✓	1 ✓	1	1	1	1 ✓
Discipline	2 ✓	1 ✓	2 ✓	2 ✓	2	2	2	2 ✓
Social support	3 ✓	2 ✓	3	3	3	3	3	3
Hours planned to study	4 ✓	3 ✓	4	4 ✓	4	4	4	4
Study choice certainty								
Prior to SA	5.0	8.0	7.0	10.0	8.0	7.0	10.0	7.0
After SA	7.0	8.0	7.0	10.0	8.0	7.0	10.0	7.0

Note.

¹ 1...4 Order of test taking from 1 (first test taken) to 4 (last test taken)

² 'high-risk' score 'medium-risk' score 'low-risk' score

³ ✓ Feedback consulted

4.3.2. Obtained scores (RQ2)

A summary of the obtained subtests scores is provided in Table 4.1. Overall, participants' scores were in the (relatively) safe areas on most subtests. One participant obtained a score in the 'high-risk area' on the numerical skills test.

4.3.3. Test taking behaviour and feedback consultation (RQ3)

A summary of the number and order of subtests taken and feedback consultation is provided in Table 4.1.

Number of subtests taken. Even though participants were instructed to be in charge of which subtests they would take and in which order, all participants completed all subtests. This is remarkable, as some participants commented that in particular the numerical skills did not seem relevant to them. Reasons for still taking this test were the few subtests in the SA:

Normally I would have skipped the numerical skills test, as I do not think it is relevant for my study decision (...). Now I filled it out, because there were not that many tests and the other tests did not consist of many questions, so I decided to see what insights the numerical skills test might provide me. (Participant P)

And the lack of clarity (despite instruction) that it was possible to skip subtests: "I thought I had to fulfil it, or I would not be able to continue with other tests" (Participant L).

Order of taking subtests. In general, participants took the tests in the order in which they were presented from top to bottom. The (incidental) reason to diverge from this order was the drive to first take the test they felt most insecure about: “Study intentions grasps my attention, as I know that, traditionally, I have the most trouble with that. That is why I am going to start with that one” (Participant P)

Feedback consultation. Two participants consulted the feedback on all subtests. Three participants did not consult any of the feedback information, as they did not notice the result button: “I really did not see the button; otherwise I would have clicked on it. I would really like to see it now” (Participant I). Though instructed about the button, apparently the button was not clear to all users.

Furthermore, three participants consulted the feedback only for some of the subtests. In those cases, feedback on social support and/or hours planned to study was neglected. These students did score relatively well on these tests, which was also mentioned as the main reason to skip the feedback: “Well, what else can I do? I ticked all the boxes (...) so I thought there is nothing to improve or do, it is fine like this and I feel comfortable with that” (Participant Y).

4.3.4. Reactions during test taking and responses to feedback (RQ4 and 5)

In this section, we discuss reactions during test taking (process aspect of validity) and how participants responded to their obtained scores and feedback (consequential aspect of validity) per subtest, before discussing these results for the SA in general.

Numerical skills – Process aspect of validity. For many participants the numerical skills test gave rise to feelings of insecurity (e.g. test-anxiety, feeling incompetent), both in advance and while taking the test. This became clear from actual statements uttered (e.g. “I will never manage this, I am so bad at mental arithmetic” (Participant L)), as well as other signals: repeatedly sighing, scrolling up and down, indicating that the test will take a long time or that by looking at how many questions still have to be filled out. For some, this test raised awareness that these skills may be important, for many the test created feelings of frustration and/or doubts about the relevance of this test. For instance for participant P, stating, “I am surprised about the math exercises, it has little to do with the study I am interested in”.

Feelings of insecurity bring forward different strategies for completing the test. One person mentioned to read extra carefully and write things down, because of finding it difficult (i.e. “Ok, fractions (...) I find that hard, so I'll have a closer look at it” (Participant I)). However, quite a few ($n = 5$), remarked that they just guessed some answers in order to complete the test. Furthermore, striking about this test was that, in contrast to the other tests, almost half of the participants felt ill at ease because the researcher was observing how they proceed through the test. Two participants even mentioned that, because of this, they filled it in at speed, at the expense of accuracy.

Numerical skills – *Consequential aspect of validity.* Although the test tended to evoke frustration, insecurities, and invalid answering strategies (hurrying, guessing), the responses on the scores and feedback were rather positive. The most common reaction was relief regarding the obtained score: “I never took math classes or anything like that, so this is not so bad” (Participant Y). Two participants had expected to score better, while four had expected to score lower than they actually did. This appeared to raise their confidence regarding their own abilities: “That is interesting, I believe I can do this” (Participant J). The feedback also resulted in reflection on the relevance of numerical skills and two participants intended to consider the possibilities for further preparation (quote 15). As participant P stated, “Apparently there is a correlation between numerical skills and obtaining study credits, I did not know that. I clicked on a link to read more about that”. One person maintained her opinion that the test was not relevant for the specific study direction she was interested in, and therefore did not recognize the added value.

One participant (L) scored in the ‘high-risk area’ on the numerical skills test. When she read the feedback, she understood that her score related to lower chances for obtaining study credits, which she mentioned as the reason for feeling a bit discouraged. Her score did not surprise her, because she always experienced problems about arithmetic, which she also expressed when taking the subtest. While reflecting on the feedback she mentioned to feel scared, though generally hopeful, because she scored well on the other tests and would not have to do that much with numerical skills in her study direction of interest, i.e. law.

Discipline – *Process aspect of validity.* In general, during this test, participants verbalized their reasoning towards an answer, for instance how they based it on previous or similar (study) situations. They also indicate to be aware that it can be hard to stay disciplined when, for example, there are other, more enjoyable, things to do. One participant said she found it difficult to answer the questions, as she had no recent or similar experiences to draw from. This test was the only test in the SA in which a possible response flaw became apparent with one participant commenting that he did not fully trust his own answers. His score was sufficient and he indicated that he tried to answer as honestly as possible, but also knows that this might turn out to be a problem.

Discipline – *Consequential aspect of validity.* One person scored lower than expected on the discipline test. This made her doubt her own answers on the test. After all, she did see herself as a disciplined person. In general, however, the discipline test results mainly reflected participants’ self-views: “Yes, of course in dark green [visualization of the score], I knew that already” (Participant J). They went through this feedback faster, compared to the feedback on the numerical skills test. One person mentioned that he merely made a quick scan with the intention to read it more carefully if the feedback would mention something surprising.

Social support – *Process aspect of validity.* For five participants the test prompted adequate reflections in regard to social support. They summarized, for instance, which persons in their environment they had already discussed support with:

My parents want to support me financially. Emotionally as well, there is lot of interest in what I do. Practically, I think so, I don't have children [*example given in the test*], but I think if I have to cancel things that people will understand that I have to study. (Participant Z)

Social support – *Consequential aspect of validity.* For one person this test was quite confronting, in the sense that it made her aware of the fact that she really has to do it on her own. For others the test was a confirmation of what they had already considered. Specifically in regard to social support, an interesting observation was that a maximum score triggered two opposite effects regarding feedback consultation. For one person, obtaining the maximum score was a reason to skip the feedback, as there is no room for improvement, whereas another person nevertheless wanted to see what the feedback said. In general, the feedback on this test evokes further reflection. For example, they think about previous studies they have done and what kind of support was helpful to them then. They also think about whether they have secured all types of support or whether they could do anything for further preparation:

I see that I am prepared quite well, I have talked to people about this. This did not happen overnight, I have weighed things and I also see that especially my husband supports me in this and we will be able to do this. (Participant I)

One participant mentioned that she does not receive all of these sources of social support, but also does not feel a need for them. Thus, her score indicated room for improvement in social support, which was not in line with her personal needs. As a result, she was confused when receiving her obtained score; she began to wonder whether she completed the test correctly.

Hours planned to study – *Process aspect of validity.* Thoughts expressed by participants while filling out this test indicate that the hours planned to study had already quite extensively been considered prior to taking the test:

I have already calculated that I have 15 hours to spend on studying. I work 2 days, so 3 days I am free and the children are at school for 5 hours then, so then I have 15 hours to study. (Participant I)

In addition, they did seem to think about the consequences of specific answers, yet that did not distract them from answering honestly: "I think I need to do more in the numbers of hours planned to study but I will stick to the 6-10 hours anyway" (Participant J).

Hours planned to study – *Consequential aspect of validity.* The obtained scores and feedback on this test mainly raised awareness of how long it will take to complete a study program, given the number of hours planned for studying. For this purpose, the feedback includes a calculation example that helps prospective students to

gain insight into how long it will take them to complete a study program, based on the number of hours they plan to study (i.e. Participant P: “This is good, an open door really, but I did not calculate it like this yet”). Although for some this means that they will spend a considerable period of time studying, it does not demotivate them: “It was a confirmation. I do like studying, so I do not really care about the nine years. It did not demotivate me, the time indication” (Participant J). For one person, the feedback did not have added value, because she already made the calculation together with a study advisor.

Overall – *Process aspect of validity* Even though all tests included in the SA are relevant in terms of ‘study preparedness’, it was not anticipated that prospective students would take all subtests. Still, participants in this study did take all subtests. Moreover – made overt by the think-aloud protocol – they seem to make an adequate translation of their personal situation and/or self-image into an answer to various test items. The numerical skills test, the only ‘cognitive’ test included, clearly evoked frustration and stress (i.e. “The stress level goes up for a little with those first questions”, Participant Y), even though most of the participants scored well on it. To some extent, this is inherent to the content of the test, yet we will have to consider how to minimise this effect, as we do not want to discourage respondents unnecessarily.

Overall – *Consequential aspect of validity* In general, it can be said that the SA provides food for thought (e.g. about social support, relevance of numerical skills) and feedback for action (e.g. calculating study time, intentions for further reading). Participants find the feedback clear and praise the headings and links, which makes it easier for them to read. However, some also indicate that they scanned through the feedback quickly and read more intently when seeing something striking.

4.3.5. Further orientation and preparation (RQ6)

Three participants reported that they are planning to take some steps for further orientation or preparation. One participant wanted to gain additional insight into the fit between her interests and a specific study direction, so she planned to discuss this with a study advisor. Two participants mentioned that they will make further inquiries regarding numerical skills, e.g. through links included in the feedback. Other participants indicated that they are not planning to take further steps in orientation. The main reason, mentioned by three participants, is that they do not think it is necessary, because they already took diverse orientation steps. Participants also indicated that it depends on the obtained score whether there is an intention to do something with the feedback:

It depends 100% on the score to what extent I am inclined to do something with it, because you do want to make it a success and if you see that one success factor is a bit less than others, you want to work on it. (Participant Y)

And they do not feel like their obtained scores indicate that they should take further action:

I would have, if something surprising resulted from that test. For instance, if discipline would have been low, should you even consider taking a study program focused on self-study? In that case, I would have liked to talk to a student, alumnus, or study advisor. (Participant E)

4.3.6. Study choice certainty (RQ7)

A summary of participants' study choice certainty is provided in Table 4.1. Most participants in the present study were rather certain already of enrolling in a course or study program at the OUNL. Study choice certainty changed only for the participant reporting a certainty of 5 prior to the SA. She was more certain of the decision to enrol afterwards (7), because her insecurity about numerical skills turned out to be unjustified and the SA raised awareness of the time it would take her to complete a study program: "It is higher than 5 now, because of the confirmation in arithmetic, that I don't have to be insecure about that, and the realization that if it takes me 9 years, I wouldn't mind so much" (Participant J).

In general, the SA did not seem to have an impact on study choice certainty. For some participants, fulfilling the SA took place after what they experienced as an elaborate orientation process. Participants stated that they believe the SA to be of more influence in the beginning of the orientation process (e.g. Participant P: "If I were still at the beginning of my orientation, then it would still have an influence. Now it is like another drop in a bucket full of water") and that the SA in itself has an impact only on study choice (certainty) as a part of a broader pallet of orientation activities. Three participants indicated that their insecurity lies mainly in the choice of study direction and the SA does not provide any tests on that. It is also noteworthy that two participants (participant Y and I) mentioned that they were planning to just start and see how they experience and perform (in) the first half year.

Though their study choice certainty did not change, five participants (both very certain and not so certain) mentioned they felt affirmed after taking the SA. Participant P, for instance, said, "The test could only have affected me negatively, but there were no big red flags to find that. Now it was more an affirmation". Participant Y stated the following:

Before I started the test, I thought I was not prepared that well and that I had not thought very well about the study I was going to do. Now I think that I actually did think well about it and I have not rushed into things. So this test may have made me even more certain that I have made the right choice.

And participant E stated, "If you still have some doubts, the test can remove them and if you are almost certain, the test can give you confidence that you are making the right choice". Three participants mentioned that it did trigger reflection on how to start well prepared:

In general, it is a good test (...) It gives you a realistic picture of how much study time you have to put in and how long it will take you and also, that it is important that you think about the financial picture and personal support, so it gives you all kinds of facets to think about. (Participant I)

4.3.7. Other validity evidences

Though the present study was targeted at process and consequence validity, the think aloud and interview data also revealed results on the *content aspect of validity* – the relationship between a test's content and the construct it is intended to measure, referring to themes, wording, and format of items on an assessment instrument (Beckman et al., 2005). In regard to the content of the SA as a whole, participants find the content relevant and understand the choices for the current set of subtests. Nonetheless, they have reservations about specific tests. Regarding the numerical skills test some indicate that they assume that this test is chosen to (partly) measure their intelligence, which they do consider relevant content for the SA. However, several indicate that they would expect another test to measure intelligence (i.e. reasoning skills) instead of or in addition to the current numerical skills test.

The tests on discipline and social support, raised doubts with three participants who thought the number of items the tests relied on was too limited to draw sound conclusions from. In addition, they commented on the formulation of specific test items, e.g. they found it hard to interpret words like 'often' (I often do not finish what I planned, because I feel lazy or tired) 'hard' (I find it hard to stick to a (study) schedule), or receiving support 'to some extent'. Finally, some participants questioned the relevance of the social support test, since it does not take into account to what extent people experience a need for various kinds of support.

4.4. Discussion and conclusion

The present study was a mixed method study aimed at investigating the process and consequential aspects of validity of a self-assessment for informed study decisions in higher online education.

Regarding the *process aspect of validity*, a general point of concern is that self-assessments, i.e. self-report measures, may be subject to all kinds of measurement errors, due to inaccurate self-perceptions (Dunning et al., 2004) or social desirable answering (Niessen et al., 2017; Viswesvaran & Ones, 1999). In the present study, one participant hinted at this stating that he did not fully trust his own answers on the discipline test. However, in general, our results demonstrate evidence in support of the process aspect of validity as the think-aloud protocol reveals that prospective students appear to base their answers on adequate (sensible) reflections. This evidence was most prominent in the non-cognitive tests (i.e. discipline, social support, and study intentions): participants brought to mind examples from their personal environment and current or previously experienced circumstances in order to decide which answer to select.

The numerical skills test specifically revealed two typical response processes, arising from feelings of uncertainty that are stirred up by the test. Most participants react on this, by adopting the strategy to fill in the test in a hurry and to guess the answers on questions they cannot answer immediately. Occasionally, this leads participants to the opposite approach: taking their time, writing down calculations and reading questions several times. Though the research context (read: the presence of an observer) may have played a role in this as well, these kind of responses are partly inherent to this type of test (Abbasi & Ghosh, 2020; Dowker et al., 2016; Liebert & Morris, 1967).

The limited number and shortness of tests in the SA appeared to motivate prospective students to take all subtests, even those that initially did not seem relevant to their study of interest. We consider this as an advantage to the process aspect of validity, as all the tests provide relevant insights independent of the study of interest (Delnoij et al., 2021).

An important threat that came to light in the current study is that some users missed the result button. Consequently, they missed important feedback information that can support them in choosing and preparing for a study in higher online education.

With respect to the *consequential aspect of validity* it appears that the SA feedback triggers reflection. The obtained scores and feedback on the numerical skills test were generally positive, in contrast to what some prospective students expected while taking the subtest. The feedback taught them that they could influence their skills by taking time and effort to practice. This resulted in enhanced self-efficacy – a person's sense of their own ability to accomplish something successfully (Bandura, 1977). We see this as an advantage for the consequences of the SA, as self-efficacy is an important determinant for students' motivation and success in higher online education (Harnett, 2016). The feedback on the other tests triggers reflection, in particular tests on social support and hours planned to study. Here, prospective students start to rethink their preparedness and intentions and whether they could do more.

However, the feedback hardly appears to influence further actions for orientation or preparation. The main reason appears to be that the prospective students in the present study had already undertaken many orientation activities. For example, they had already spoken with a study advisor (which is also recommended in the feedback on the SA), they attended an open day or orientation day of a specific study direction and consulted the information on the website. In addition, they indicated that, to them, their scores did not imply that further preparation was necessary and that they might have followed up on the feedback more if their scores had been lower.

Furthermore, the SA did not appear to have a big impact on study choice certainty. This finding must, again, be interpreted against the same background of a relatively well-prepared group of participants who felt already quite certain before completing the SA. None of the participants felt less certain or discouraged, but of course, their relatively high scores gave no reason for this. In general, participants in the present study stated that the SA would have had a bigger impact with respect to following up on

the feedback and/or study choice certainty if they had taken it earlier in their study orientation process. This explains why many of the participants indicated beforehand that they were mainly looking for affirmation. In that sense, the SA did meet their expectations. Overall, these results appear to be in line with other research. For instance, Soppe et al. (2019) have already shown that study choice certainty plays an overarching and important role in (the absence of) the effects of various study orientation activities. They also have demonstrated that the more certain prospective students are about their initial choice, the less impact an orientation activity has on their final choice and, thus, the less likely a change in choice certainty will take place. An interesting finding in their study was that some participants, who were 100% certain initially, nevertheless said that the orientation activity made them even more certain. So, it seems that affirmation is an important consequence even for those who may not appear to need it.

4.4.1. Implications for the SA, theory and practice

Implications for the SA. For the current SA specifically, based on the present study, some refinements are proposed, before 'mainstream deployment'. First, recommendations are based on the evidence and threats in regard to the SA's *content*, despite the current study's focus on process and consequential aspect of validity. Results indicate that an addition of test items to the discipline and the social support test as well as an addition to the present set of subtests should be considered to reduce the threat of construct under-representation (Downing & Haladyna, 2004). Regarding additional items to existing subtests, further analyses should be carried out to secure the internal structure and predictive value of the tests. At the same time, when adding test items or subtests to the SA, parsimony should not be lost sight of, as the limited number and shortness of tests did motivate students to take all subtests, even those that did not seem relevant to them initially. In regard to adding new subtests, a broader range of knowledge and skills tests would be valuable (e.g. reasoning skills, study strategies) and a content sample test would be recommendable. After all, prospective students indicate they expect and desire some feedback regarding the fit with the subject of study they are considering to choose. A content sample subtest can offer them a hands-on experience prior to enrolment. Ideally, this would consist of for instance, studying course literature and/or watching video-lectures, followed by a short exam (Niessen et al., 2018).

Secondly, results in regard to the *process aspect of validity* showed that the numerical skills test seems to create a stressful state of mind regarding the SA that eases in the other tests with questions that merely require an answer realistically reflecting personal characteristics or circumstances rather than a correct answer. Since prospective students seem to fill out the SA from top to bottom, it is recommended either to change the linear presentation of the subtests or to change the order of the tests so that the numerical test is not the first test they encounter. In general, the SA should not frustrate or discourage students more than necessary. In that respect, we recommend to monitor test-anxiety and avoidant test-taking strategies in further evaluation as well.

A final refinement for the SA concerns the result button. To prevent prospective students from missing out on relevant feedback information, it is suggested to consider

a push communication strategy (e.g. an automatic pop-up feedback window after taking a test) instead of the current pull strategy. In that way, no extra attention is required from users by which they are more likely to take the feedback in and perhaps act on it.

Implications for theory. More generally, this study adds to the literature by providing a distinctive and authentic example of collecting and interpreting process and consequential evidence with the aim to enhance assessment validity. Though validity literature provides a clear picture of the different sources of evidence and threats to validity, a flaw of many applied validation studies is that they tend to focus solely on content, internal structure and predictive aspects of validity (Cook et al., 2014). Moreover, regrettably these examples mainly involve so called high-stakes assessments (i.e. for selection, pass/fail, or grading decisions), standardized tests, predominantly in the context of health professions (Cook et al., 2014). As our results showed, a self-assessment can have an impact in prospective students' study decisions and progress. Access to higher education – even if (or especially when) it is open – requires the best possible decision making support. It is a call of duty to justify assessment procedures in this context, based on empirical arguments (Niessen & Meijer, 2017).

Implications for practice. The self-assessment is embedded in the existing practice of providing information and advice prior to enrolment. Combining orientation activities with expert advice has been shown to be relevant for the quality of study decisions and the study process (Borghans et al., 2015; Zhang et al., 2019). Hence, study advisors were closely involved in the development process of the SA and especially of the feedback provided aligned to the subtests, as this feedback refers to study advisors' services. Based on this feedback, prospective students, thus, might contact study advisors for further clarification or advice in following up the feedback. This assumes that study advisors are able to interpret the SA results with the necessary nuances. In that regard, recommended future steps involve additional training (e.g. a handout of how to interpret SA scores) and exchange of experiences, for quality assurance purposes.

The SA evokes reflection on study preparedness and offers concrete insights and suggestions regarding opportunities to improve chances of success, both prior to and after enrolment. The 'advice' category in the feedback links for example also to existing remedial tutorials and courses the educational institute provides to its students. Previous research has shown that such (early) remediation is a promising effective strategy for improving retention (Delnoij et al., 2020a; Muljana & Luo, 2019; Robinson et al., 1996; Sage et al., 2018; Wachen et al., 2016).

4.4.2. Limitations and implications for future research

Reflecting on the specific research method used for this study, an observer effect (i.e. the Hawthorne effect, see Sommer, 1968; or McCambridge et al., 2014 for a more recent review) might have played a role as the researcher was watching participants while taking the test. For instance, regarding the numerical skills test, some participants mentioned that they felt rushed or insecure, because of being observed. In general, however, there were only few indications of flawed answers. Some participants indicated the tendency to choose a specific answer option because that might lead to a higher score, but

eventually selected their original answer. Still, the results have to be interpreted with some caution.

For future research, we recommend to expand the investigation of the consequential aspect of validity by evaluating the effects of the SA on enrolment and study success after enrolment (Downing, 2003). In that regard, the classification model (i.e. accuracy, false positives/negatives) set in an earlier stage of the design process (Delnoij et al., 2021) should be evaluated. In addition, the current sample involved a relatively large group already reasonably certain of their study decision while participating. In the present study, the sample consisted of prospective students who indicated their interest by, for instance, calling the student service office (see method section). It seems that students do so, in case they are already relatively certain of enrolling. Future research is needed to investigate the SA's impact on prospective students who are less certain of their study decision (Cobern & Adams, 2020; Guba, 1981; Shenton, 2004). In that regard, we recommend utilizing an additional or different sampling method.

Nevertheless, relatively rapid and innocuous pilot tests like the present study are important in design-based research in general and for the SA in specific, to enable adjustments and refinements aligned to the intended effects prior to a 'full release'. In addition, small-scale qualitative studies provide in-depth insight into prospective students' response processes while taking the SA and the consequences of the SA on their study decision process, two aspects that are underreported in applied validation studies, yet tremendously important in determining assessment effectiveness.



Chapter 5

- Do self-assessments for informed study decisions actually inform study decisions? A model for evaluating the consequential validity aspect

This chapter is based on: Delnoij, L. E. C., Janssen, J. P. W., Dirkx, K. J. H., Vogten, H., Martens, H., Elston, S., Hermans, H., & Martens, R. L. (2021) *Do self-assessments for informed study decisions actually inform study decisions? A model for evaluating the consequential validity aspect*. [Manuscript submitted for publication]. Faculty of Educational Sciences, Open Universiteit.

Self-assessments prior to student enrolment are a promising way to address student commitment and retention in an early stage. Such assessments aim to inform study decisions by evoking reflection on study preparedness and providing advice for further preparation. These assessments require a solid and extensive validation process. Validations reported in the literature so far tend to ignore the consequential aspect of validity: assessment impact and fairness. The current explanatory correlational study addresses this gap and sets an example. Prospective students ($N=662$) orienting towards studying in higher online education took a self-assessment consisting of a variety of subtests. The self-assessment's impact appeared in line with its purpose for a reasonably large group of prospective students (68.9%). Their study choice certainty was adapted or remained unchanged in accordance with their obtained scores. Moreover, study choice certainty after taking the self-assessment was positively related to enrolment probability. In addition, the impact of the assessment was fair (similar across subgroups), although men's study choice certainty appeared relatively robust against unfavourable scores. Implications for developing self-assessments for informed study decisions are discussed.

5.1. Introduction

In order to enhance retention in higher (online) education, it is advised to address student commitment in an early stage, even prior to enrolment (Muljana & Luo, 2019). One way to do so is by providing self-assessments prior to student enrolment, in which prospective students receive information about where they stand in regard to the demands of studying in higher (online) education (e.g. Nolden et al., 2019). These instruments are advisory and information instruments, which are conducive to self-examination (Hornke et al., 2013). In such self-assessments, prospective students complete tests on factors that are proven relevant for a solid start and continuation in higher education (Delnoij et al., 2021; Nolden et al., 2019). For instance, in the Self-Reflection Tool by Nolden et al. (2019), prospective students can take subtests on factors such as self-discipline, motivation, and learning strategies. Feedback aligned to those tests is aimed at raising awareness and self-reflection (Nolden et al., 2019; Broos et al., 2018; 2019; Delnoij et al., 2020). Feedback to prospective students as presented by Broos et al. (2018; 2019), for example, involves information on the obtained subtest scores, compared to successful students in the first year after enrolment, and advice for further preparation. Such information is presented to support prospective students in making well-informed study decisions and possibly leads to early remediation, all for a successful start and success in higher education (Broos et al., 2018; 2019; Kubinger, et al., 2012; Muljana & Luo, 2019; Nicol, 2009; Nolden et al., 2019; O'Regan et al., 2016; Tinto, 1999; Van Klaveren et al., 2019).

With an eye on the possible continuation of online education after the Covid-19 pandemic (Gomez Recio & Colella, 2020), strategies to enhance retention in the context of *online* higher education have become even more relevant. There is no doubt that retention is a serious issue in this context, as demonstrated by its place on institutional agenda's for many years (Muljana & Luo, 2019; Rovai, 2003; Simpson, 2010). Considering that self-assessments prior to student enrolment seem a promising approach in higher education more generally (Fonteyne & Duyck, 2015; Kubinger et al., 2012; Lee et al., 2013; Muljana & Luo, 2019; Nolden et al., 2019), we developed a similar instrument for prospective students in higher *online* education. So far, the development and validation of these types of self-assessments has received little attention, or, at least, has not been reported on publicly (Niessen & Meijer, 2017). In other words, there is a lack of clarity about when and under what circumstances such instruments are effective. Therefore, this study aims to provide an example of how the impact of self-assessments can be evaluated to shed light on the validity of decisions based on these assessments. More specifically the consequential aspect of validity, e.g. the impact on enrolment decisions.

Based on previous studies (Delnoij et al., 2020a; 2020b; 2021), the self-assessment (hereafter 'SA') evaluated in this study entails six subtests divided in three categories of tests: *attitude* (consists of subtests on discipline and study expectations), *knowledge and skills* (includes subtests on numerical skills and study strategy use), and *personal situation* (entails subtests on social support and hours planned to study). Prospective students choose which and how many subtests they want to take. On each subtest, they receive feedback consisting of information about their obtained score in relation to the chances of success after enrolment (i.e. a 'high', 'medium', or 'low-risk' score). An explanation of

the measurement and concrete suggestions on how to address possible risks indicated by their obtained score complete the feedback information (see Delnoij et al., 2020b). The SA does not differentiate between or provide advice for specific study directions. Also, it is not committal or aimed at selecting students. Rather, the aim is to enable informed decision-making (food for thought), and to encourage prospective students to start well-prepared (feedback for action).

5.1.1. *The quest for validity*

Instruments providing prospective students information about their possible future success potentially have far-reaching impact both for the individual (student) in terms of decision-making and progress, and for the institute in terms of enrolment and success rates. For instance, a study by Van Klaveren et al. (2019), showed that providing students with feedback on expected success rates increased enrolment with about 25%, but did not reduce first year dropout. It is important that such an impact is theory- and data-driven (Nolden et al., 2019). Therefore, assessment procedures in the context of access to higher education - even for non-selective purposes - require a substantive and solid development and validation process (Niessen & Meijer, 2017).

Hence, in line with modern validity theories, the development process of the SA at stake in the present study involves evaluation of five sources of validity evidence (AERA et al., 2014), corresponding to the *content*, *predictive*, *internal structure*, *process*, and *consequential* aspects of validity. Table 5.1. explains all five sources of validity evidence illustrated with examples in the context of self-assessments prior to student enrolment.

Table 5.1. *Explanation and examples of the five sources of validity evidence*

Validity aspect	Explanation	Example in SA context
Content	Extent to which the test content accurately represents the content domain.	Self-assessment contains subtests on variables sensitive to change or remediation, relevant to study success (domain), in the context for which the self-assessment is developed.
Predictive	Relative performance of test scores in predicting (supposedly) related variables.	The accuracy with which the self-assessment scores predict success after enrolment.
Internal structure	The degree to which items reflect coherent dimensionality, both on theoretical and statistical grounds.	Dimensionality of subtests and degree to which subtest items correlate.
Process	The fit between what the items/test intend to measure and the experiences and considerations test takers appear	Actions, strategies and thought processes of prospective students taking the self-assessment, e.g. to what extent do answers appear to be based on sound and realistic reflections on, e.g. actual behaviour or

	to take into account when responding.	experiences so far. Includes possible influences such as social desirable answering, test-anxiety, etc.
Consequential	Interpretations of and actions following test result and feedback are correct and in line with intended uses.	Extent to which self-assessment affects enrolment decision.

Note. Explanations are based on AERA et al. (2014), Beckman et al. (2005), and Cook et al., (2014).

Investigating these sources of validity evidence is an iterative and ongoing process, as student populations and/or educational practice evolve over time (Messick, 1990; Royal, 2017). However, some chronological order appears to exist when it comes to collecting evidence from these sources. For instance, investigating response processes and consequences makes sense only after the content, internal structure, and predictive power have been more or less assured.

What is striking when looking at applied validation studies, however, is that they tend to focus mainly on content, internal structure and predictive power and that there is less attention for the process and consequential aspects of validity (Cook et al., 2014; Kreiter, 2016). Cook et al. especially argue that greater emphasis is required on describing and defending the use of scores and the decisions and actions following score interpretation, i.e. on the consequential aspect of validity. In the context of study decision support tools, it appears such empirical evaluation is often implicit or even completely lacking (Niessen & Meijer, 2017).

After having established satisfactory results on other aspects of validity in previous studies (Delnoij et al., 2020a; 2020b; 2021; in review), this evaluation study aims to gain insight into evidence of consequential validity of an online SA for informed study decisions in higher online education. In doing so, we not only aim to assure the validity of this particular instrument but also to fill the gap in the literature and to set a standard for validating such orientation instruments. The procedure we show in this paper may inspire other institutions regarding the design and decision process for assessment instruments aimed at informed study decisions (Neumann et al., 2020). Next to that, this study provides insight into the impact of such instruments, which is a relatively underexplored field (Niessen & Meijer, 2017). In the next section, we dive deeper into the concept of consequential validity, before we present the specific research questions of this study.

5.1.2. Consequential validity: What it is and how it can be evaluated

The consequential aspect of validity regards the intended and unintended impact of the assessment, both positive and negative, for the individual and/or society, i.e. the soundness of decisions made and actions taken, based on assessment results (e.g. taking a remediation course to address sub-standard performance)(AERA et al., 2014; Beckman

et al., 2005; Cook et al., 2014). Validity literature highlights three themes with respect to the consequential aspect of validity: *pass/fail cut-off establishment and consequences*, *individual and societal impact of assessment scores*, and *fairness* (AERA et al., 2014; Beckman et al., 2015; Cook et al., 2014; Downing, 2003; Messick, 1995; Niessen, 2018; Schreurs, 2020; St-Onge et al., 2017). We briefly discuss each of these themes and illustrate them with examples in the context of self-assessment for informed study decisions.

Pass/fail cut-off establishment and consequences. A first important concern when it comes to the consequential aspect of validity is the establishment and impact of score distributions. The score distribution provides context and meaning to all possible scores on a test, e.g. what is the range of ‘sufficient’ scores. In the context of self-assessments, the score distribution determines the cut-off point below which for instance remediation is considered beneficial (Cook et al., 2014). This score distribution of subtests must be substantiated and documented (Downing, 2003). A score distribution can be established in various ways. Nolden et al. (2019), for instance, created three categories of scores based on the means and standard deviations of successful students and non-successful students. In their score distribution, a ‘risk’ score (or ‘red zone’ as they call it) entails all scores lower than the mean score of non-successful students minus one standard deviation on a particular test, while a ‘safe’ score (‘green zone’) entails all scores higher than one standard deviation above successful students’ mean score. Scores in between these ‘extremes’ constitute the ‘yellow zone’. Another approach, and one we followed for the SA, is to take classification accuracy into account in determining the score distributions. Classification accuracy is a metric indicative of the performance of a variable (read: test score) in relation to a classification model (i.e. classifying completers and non-completers) and is calculated by dividing the number of correct predictions (‘true positives’ and ‘true negatives’) by the total number of predictions. This approach enables to reckon with the fact that, in this case of open online education, we want to be particularly careful not to unnecessarily discourage students. Specifically, we want to minimise the likelihood of false negatives to a maximum of 5%, even if this means a trade-off in terms of the proportion of correctly identified actual non-completers, which a previous study established at 13% (Delnoij et al., 2021). Based on those results we determined the cut-off scores for the subtests. For instance, prospective students receive a ‘high-risk’ score when scoring lower than 30% on the discipline test, a ‘medium-risk’ score when scoring between 30 and 60% and a ‘low-risk’ score when scoring higher than 60%. More specifically, this means that a maximum of 5% of students receiving a ‘high-risk’ score might in fact appear completers.

Continuous monitoring of the classification model and score distributions is an important part of evaluating the consequential aspect of validity (Cook et al., 2014; Downing, 2003). The classification model of the SA is based on whether or not any study credits were obtained within a year after enrolment (see Delnoij et al., 2021, for a detailed description). Hence, monitoring the classification model requires longitudinal evaluation on the same outcome measure. Moreover, for prospective students who decide not to enrol (either following a favourable or unfavourable score) it is impossible to establish whether the prediction appeared accurate. This means that purely theoretical, evaluation

of the classification model is necessarily restricted to a subset of test takers, i.e. those who decide to enrol.

Individual and societal impact. A second theme addressed in the literature on the consequential aspect of validity is the individual and societal impact of an assessment instrument, already implied in the previous section when referring to possible discouragement of prospective students, but meriting further elaboration. In the context of the SA investigated in this study, possible individual consequences involve various more or less successive variables: *interpretation* of obtained scores and feedback, intentions and actual steps in *following-up on the feedback*, *study choice certainty* (self-confidence), the *decision* to (not) enrol, and finally, *achievements* after possible enrolment. Both interpretation of scores and feedback and intention to follow-up on feedback have been addressed in a previous study (Delnoij et al., in review). The present study's focus is on study choice certainty and enrolment decisions. Consequences in terms of post-enrolment achievements require additional longitudinal investigation at a later point in time, and will be, as mentioned, necessarily limited to a subset of test takers, i.e. those who end up enrolling. Gaining insight into the impact of the SA on the enrolment decision and possible factors playing a role in this is essential in itself as well as an indispensable part of understanding the full picture.

The extent to which individual consequences can be considered *valid*, in turn, depends on the wider picture, i.e. what went before, and possibly after. Increased levels of study choice certainty, for example, can be considered a valid consequence for those who score well on the SA. As can feeling affirmed in an already certain choice (Delnoij et al., in review; Soppe et al., 2019). However, important to note is that scoring well on the SA or increased levels of study choice certainty do not necessarily imply that one 'has to' enrol, i.e. enrolment as the single valid consequence following next. The SA is likely to be part of a rich palette of orientation activities one can undertake and though the SA scores might imply one is 'good to go', one might opt for different opportunities discovered during the orientation. Nevertheless, a valid consequence to a risk score would be to take remedial measures as a follow up on the feedback or even to postpone or reconsider the study decision (Broos et al., 2019).

Research shows that orientation activities such as the self-assessment of the present study might affect study choice certainty. Moreover, these studies (Soppe et al., 2019; Delnoij et al., in review) emphasize that the impact of such orientation activities on study choice certainty depends on the initial study choice certainty, i.e. the level of study choice certainty at the start of these activities. Those already relatively certain about their decision, appear to be less affected by orientation activities and seem to be merely looking for affirmation of the decision they already made. These results are in line with a robust psychological effect known as *confirmation bias* (Rabin & Schrag, 1999), implying that students place more weight on signals that confirm their beliefs than on disconfirming signals (Eil & Rao, 2011). In sum, orientation activities are expected to influence study choice certainty, but this relation is moderated by initial study choice

certainty (see Figure 5.1.). Naturally, initial study choice certainty will also directly link to study choice certainty following orientation activities.

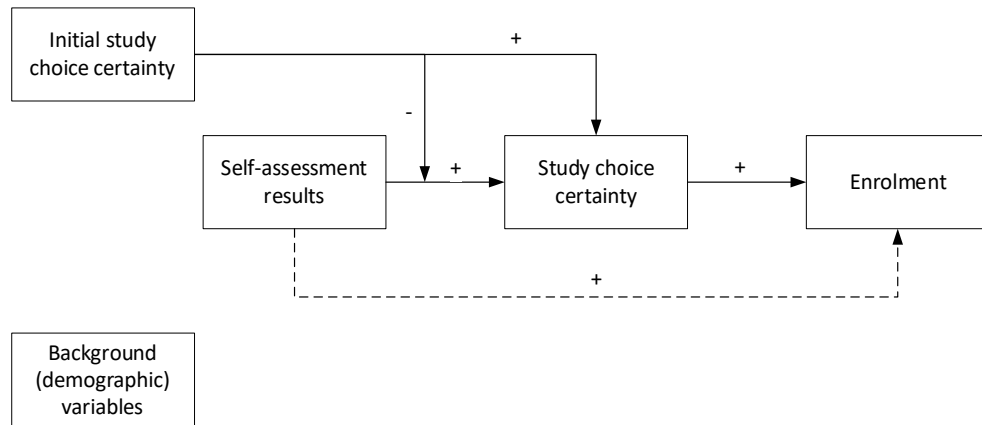


Figure 5.1. *Conceptual model*

Moving to the societal level, the anticipated consequence is a positive impact of the SA on retention, either through improved preparation by prospective students informed by the feedback in the SA, or through reconsideration of enrolment by truly at risk prospective students. A cost-benefit analysis at the institutional/societal level, requires longitudinal research at a different level of aggregation and should include costs to develop, to provide and to maintain (i.e. continued validation) of the instrument (Kraft, 2020; Schreurs et al., 2018).

Fairness. A final theme highlighted in the literature regarding concerns the fairness of an assessment instrument (AERA et al., 2014; Nisbet, 2019; Xi, 2010). In the Standards for Educational and Psychological Testing (AERA et al., 2014), fairness has been conceived as the absence of bias, equitable treatment of all test takers in the testing process, and equity in the opportunity to learn the material in an achievement test. As Kane (2010) has described, validity and fairness are closely related concepts: both focus on the interpretations and uses of test scores and whether that is appropriate for a specific target group under a range of circumstances (Kane, 2010). The relation between both concepts is perhaps best illustrated by Xi's (2010) definition of fairness as "comparable validity for identifiable and relevant groups across all stages of assessment, from assessment conceptualization to the use of assessment results." (p. 154).

However, so far, there is only limited research on fairness in the ultimate score interpretation and score-based decisions (i.e. the consequential aspect of testing) in the assessment literature (Kreiter, 2016; Xi, 2010). With respect to self-assessments for informed study decisions in higher education, fairness plays an important role. In the context of open (i.e. non-selective) education, we want to be particularly careful not to unnecessarily discourage prospective students by the feedback provided in the SA. We

aim to provide prospective students a realistic perspective on their preparedness for enrolment in higher online education and how this might be strengthened in case the test scores suggest such actions. Yet, the feedback provided should not have a different impact on study choice certainty of test takers with similar scores, but different backgrounds: the impact of a risk score on the discipline test should not be different for someone with a lower level of prior education than for one with a higher level of prior education. As visualized in Figure 5.1., background variables should not moderate the relationships between self-assessment, study choice certainty and enrolment.

In the present study we include fairness in our investigation of the consequential aspect of validity not only in terms of the impact on study choice certainty but also in terms of the consistency of score-based decisions (i.e. enrolment) for different groups, based on demographic variables such as gender or prior level of education.

5.1.3. Research questions and hypotheses

All in all, this study aims to evaluate consequence evidence of an online self-assessment for informed study decisions in higher online education. More specifically, the study addresses the impact of the SA on a) study choice certainty (i.e. impact on an individual level), b) the decision to enrol (i.e. impact on both individual and institutional/societal level), and c) the extent to which any impact is influenced by specific background characteristics (i.e. fairness). To that end, the following main research question has been formulated:

To what extent do SA results affect study choice certainty, as well as a decision to enrol, and to what extent does this appear to be moderated by specific background characteristics?

In order to answer the main research question and based on the theoretical framework, the following hypotheses were formulated:

1. Prospective students who obtain favourable self-assessment results are more certain about enrolment and more likely to enrol.
2. The impact of self-assessment results on study choice certainty depends on initial study choice certainty: those students who are (rather) certain about enrolling are more likely to remain unperturbed by less favourable self-assessment results, than those (rather) uncertain about enrolling.
3. The impact of self-assessment results on enrolment is mediated by (initial) study choice certainty.
4. In regard to fairness, the relationships in hypotheses 1 and 2 are independent of background variables (i.e. gender, age, and prior level of education).

5.2. Methods

5.2.1. Design

The present study represents a particular cycle in the design-based research approach, typically comprising iterative stages of analyses, design, development, and evaluation (Van den Akker et al., 2013). More particularly, this study is part of the evaluation stage of the SA that was designed and developed for prospective students of the Open University of the Netherlands (OUNL). OUNL provides academic courses as well as full bachelor and master programs, in online, blended, and hybrid modes of delivery. The open access policy of OUNL means that the only entry requirement is a minimum age of 18 years (though, additional entry requirements may be formulated for more advanced courses).

The evaluation stage involves an explanatory correlational design (Creswell, 2014). Data were gathered through the subtests and a survey which participants were asked to fill out after completing the SA. Data include obtained self-assessment results, background variables, and (initial) study choice certainty (see section 5.2.3.). Data on enrolment behaviour was obtained via the student information system.

5.2.2. Participants

Sampling took place from December 2020 until May 2021. Prospective students consulting the OUNL website were presented the option of taking the online SA. Taking the SA, did require a registration, which involved personal data to be used for the identification of possible subsequent enrolment. Table 5.2. provides an overview of the number of participants.

Table 5.2. *Participants*

Category	Number of participants	% Of previous category
Accessed the SA	1838	N/A
Accepted informed consent	1536	83.6
SA users ¹	662	43.1
Survey respondents	231	34.9 ²

Note. ¹ Prospective students who completed at least 1 subtest in the SA; ² 15.0% of those who accepted informed consent.

5.2.3. Materials and measurement

Self-assessment. The SA, illustrated in Figure 5.2., consists of six tests: discipline, expectations, study strategies, numerical skills, social support, and hours planned to study (Delnoij et al., 2020a; 2021). Four tests (discipline, numerical skills, social support, and hours planned to study) result in a score and related feedback upon completion. Two of these tests (study strategy use and study expectations) are still prototypical, meaning that the feedback provided for these tests is not yet personalised as the predictive value and

thus, score cut offs of these tests, are still under investigation. For all subtests, an example item is provided in Appendix A. Feedback design is based on related work in various other contexts (Broos et al., 2018; 2019; Fonteyne & Duyck, 2015; Jivet et al., 2020; Nolden et al., 2019) and further informed by two user experience studies (Delnoij et al., 2020b; in review). The feedback consists of three components: information on the obtained score (Figure 5.2.F), information on the test (what was measured and why) (Figure 5.2.G), and advice for further preparation (Figure 5.2.H). Information on the obtained score entails a visualization in which the obtained score (indicated by an arrow) is projected on a bar representing the possible range of scores (scale of 0 – 100%) as illustrated in Figure 5.2. The colour in the bar fades from white ('high-risk' area) via light green ('medium-risk' area) to dark green ('low-risk' area) indicating increased odds of course completion following enrolment. Additional feedback in line with the obtained score is presented in a pop-up and can be further accessed through the overall self-assessment dashboard.

Separate variables are defined to indicate overall self-assessment results and subtest results, which are operationalized as follows:

- **Overall self-assessment result:** a dichotomous variable indicating whether prospective students obtained any 'high-risk' scores or not.
- **Personalised subtest results:** the scores on the four subtests resulting in a personalised score ('high', 'medium', or 'low-risk' score).
- **Expectations/study strategy subtest taken:** dichotomous variable indicating whether or not these tests were taken. These latter tests did not result in a personalised score (yet), but their general feedback might still affect study choice certainty.



Figure 5.2. *Self-assessment design*

Survey – Background variables. Background variables in the current study are gender, prior level of education, and age. These variables were measured by multiple-choice (gender and prior level of education) or open ended (age) questions in the SA.

Survey – Study Choice Certainty (SCC). Initial SCC was measured retrospectively by asking participants to rate their certainty of enrolling at the OUNL *prior* to taking the SA retrospectively on a 4-point scale (1= completely uncertain, 2 = rather uncertain, 3 = rather certain, 4 = completely certain). Next, they were asked to indicate their current certainty of enrolling (*after* taking the SA). For analyses, we take into SCC *after* taking the SA, while controlling for initial SCC.

Enrolment. Enrolment data was obtained via the student administration system, with a score of 1 being assigned to those who enrolled in a course at the OUNL within 3 months after taking the SA and a score of 0 to all others. We opted for a limited interval for obtaining the outcome measure, so that a possible relationship between enrolment and the SA is still plausible. The choice for a 3 months interval specifically is supported by the fact that a majority (62%) of newly enrolled students enrolled within 3 months after their first orientation experience (Expertise Centrum Onderwijs (ECO), 2021).

5.2.4. Procedure

Research procedure. Ethical approval for the ongoing research was obtained from the institutional committee of ethical (cETO) (approval code U202008923). Upon accessing the SA page, prospective students were invited to take part in this study by means of a pop-up including an online information letter and consent form. Consent was entirely voluntarily: those who declined could still take the SA, without their data being used for research purposes. In the general introduction on the SA dashboard (see Figure 5.2., part B), participants were invited to take as many tests of the SA as they liked and were asked to fill out the survey afterwards. To obtain a score and aligned feedback, all items in a subtest must be filled out. The survey could be filled out leaving any of the questions unanswered.

Analysis. All analysis were conducted in Jamovi 1.1.8.0. (R Core Team, 2018; The Jamovi Project, 2019). Descriptives are analysed regarding demographics, subtests taken (taking into account the number of available tests), and obtained subtest scores, in order to provide context for the interpretation of further results. We also tested assumptions for parametrical testing. If not stated otherwise in the results section, those assumptions were met and parametrical analyses were conducted.

All hypotheses were tested by means of regression analyses. For the first hypothesis, we analysed the relationship between self-assessment results (i.e. on an overall level (obtaining 'high-risk' scores or not) and subtest level (subtest scores)) and study choice certainty in linear regression models, while controlling for initial study choice certainty. Since all participants completing the study strategies test, also appeared to have completed the expectations test we could not include both of them separately,

as this would result in multi-collinearity. Thus, we included the test completed most often, the expectations test. In analyses on subtest level, we added all remaining predictors (and covariates) simultaneously and excluded the non-significant subtests in a stepwise backwards method, beginning with the least significant predictor. The relationships with enrolment as outcome measure were analysed in binary regression models.

To test the second hypothesis, we checked for interaction effects between initial study choice certainty and self-assessment results in predicting study choice certainty. Significant interactions will be reported in the results section.

To test the mediation hypothesis (3), we added study choice certainty to the resulting model from testing hypothesis 1 (after having established a relationship between study choice certainty and enrolment).

To test the fourth hypothesis, on fairness, we checked for interactions with background variables. For example, impact of subtest scores on study choice certainty should be equal for different genders. This means that no significant interaction effect should be found between gender and subtests score in predicting study choice certainty. Interaction effects with background variables were analysed only for subgroups with $n > 5$. Significant interactions will be reported in the results section.

5.3. Results

5.3.1. Demographics

The average age of SA users ($N = 662$) is about 35 years ($M = 34.7$, $SD = 12.0$) and 57.9% are women. Most participants (55.2%) already hold a degree in higher education (i.e. university of applied sciences or scientific university).

5.3.2. Subtests and obtained scores

A total of 475 users (72%) consulted the assessment when four tests were available, and on average took 2.80 subtests. In addition, 180 users (28%) consulted the assessment when six tests were available, taking on average 3.30 subtests. Table 5.3. provides an indication of the relative 'popularity' of the various subtests in both scenario's, as well as the overall scores on subtests. For more detailed information (correlations between subtests scores), we refer to Appendix B.

Looking at specific subtests, 'high-risk' scores are obtained most often on the social support test (8.8%), followed by the numerical skills test (5.9%). Considering the context of online (distance) education it is somewhat striking that the discipline subtest appears to be the least problematic overall. Though, this test is completed by almost all participants (which might indicate discipline is a 'general concern'), taking this subtest is most likely to result in a 'low-risk' score (70.2%). Overall, 13.4% ($n = 89$) of the SA users obtained at least one 'high-risk' score (not included in Table 5.3.). The fact that this group is relatively small is not unexpected; a rather restrictive limit has been set in the classification model for assigning 'high-risk' scores, in line with the open accessibility of the institution (Delnoij et al., 2021).

Table 5.3. *Number of subtests taken and obtained scores*

Test	n taking the subtest when		Total n (%) ³ taking the test	High- risk %	Medium- risk %	Low- risk %
	4 subtests available (%) ¹ (n = 475)	6 tests available (%) ² (n = 180)				
Discipline	464 (97.7)	171 (95.0)	635 (95.9)	4.4	25.4	70.2
Numerical skills	306 (64.4)	102 (56.7)	408 (61.6)	5.9	52.2	41.9
Social support	285 (60.0)	77 (42.8)	362 (54.7)	8.8	65.5	25.7
Hours planned to study	277 (58.3)	82 (45.6)	359 (54.2)	4.9	58.5	36.8
Study strategies	N/A	78 (43.3)	78 (11.8)			
Expectations	N/A	108 (60.0)	108 (16.3)			
Average number of tests taken (SD)	2.80 (1.31)	3.30 (1.98)				

Note. ¹ percentages based on total n taking all the subtests they took when there were 4 subtests available; ² percentages based on total n taking all the subtests they took when there were 6 subtests available; ³ percentages based on SA users (N = 662).

5.3.3. 'High-risk' scores, study choice certainty, and enrolment

According to the hypotheses formulated in section 5.1.3., on an overall level, we expect prospective students who obtained at least one 'high-risk' score to be less certain of enrolment after the SA (for those initially (rather) uncertain) as well as less likely to enrol in a course, independent of background variables.

Study Choice Certainty (SCC). Before testing the hypotheses, Table 5.4. first provides insight into the change in SCC of prospective students obtaining one or more 'high-risk' scores compared to those who did not such scores. Both a decrease⁴ in SCC for those who obtained a 'high-risk' score and an increase⁵ in SCC for those obtaining no 'high-risk' score constitutes evidence of consequential validity, whereas the opposite would indicate a threat to consequential validity. A grey filling in cells is applied to 'tentatively' indicate where results provide a threat to the consequential aspect of validity.

In general, for those not obtaining any 'high-risk' scores, SCC largely remains the same or increases, in line with expectations. However, there also seems to be a small number (n = 8, darker grey filling in Table 5.4.) who appear to become less certain about enrolling, despite not obtaining any 'high-risk' scores. The two additional tests (expectations and study strategies) might play an explanatory role here. Prospective students do not obtain personalised scores on these tests, yet the general feedback might still indicate misaligned expectations, possibly affecting their study choice certainty. This does not appear to be a plausible explanation, as only two of the 'unexpectedly discouraged' participants took these tests. Another explanation might be that these prospective students obtain relatively more 'medium-risk' scores. This would

⁴ Or staying equally uncertain (taking into account a floor-effect).

⁵ Or staying equally certain (taking into account a ceiling-effect).

indicate that, in general, they do not score very well on the self-assessment, though without obtaining 'high-risk' scores. This appears to be a more fitting explanation, as those who are 'unexpectedly discouraged' appear to (proportionally) obtain significantly more 'medium-risk' scores (Mdn = 0.63) compared to those also not obtaining 'high-risk' scores and not discouraged (Mdn = 0.26) ($U = 75.0, p < 0.001$, Cohen's $d = 1.74$).

About half of those obtaining at least one 'high-risk' score, appear to reflect adequately on their initial certainty as 18.4% becomes less certain of enrolling and 31.6% stays equally uncertain. We must note, however, that also some become more certain, despite obtaining 'high-risk' scores, 18.4% ($n = 7$, darker grey filling in Table 5.4.).

Table 5.4. *Impact of obtaining high-risk scores (no/yes) on study choice certainty*

Obtaining 'high-risk' scores	Certainty about enrolling <i>prior</i> to SA	Certainty about enrolling <i>after</i> the SA, n [%] ¹				<i>N</i>
		Completely uncertain	Rather uncertain	Rather certain	Completely certain	
No	Completely uncertain	8 [4.1]	8 [4.1]	5 [2.6]	0 [0.0]	21
	Rather uncertain	1 [0.5]	37 [19.2]	15 [7.8]	4 [2.1]	57
	Rather certain	0 [0.0]	2 [1.0]	62 [32.1]	8 [4.1]	72
	Completely certain	0 [0.0]	0 [0.0]	5 [2.6]	38 [19.7]	43
	<i>Subtotal</i>					193
Yes	Completely uncertain	4 [10.5]	0 [0.0]	0 [0.0]	0 [0.0]	4
	Rather uncertain	1 [2.6]	8 [21.1]	4 [10.5]	1 [2.6]	14
	Rather certain	0 [0.0]	3 [7.9]	9 [23.7]	2 [5.3]	14
	Completely certain	0 [0.0]	0 [0.0]	3 [7.9]	3 [7.9]	6
	<i>Subtotal</i>					38

Note. ¹ Percentages based on subtotals; 33.9% of those not obtaining any high-risk scores filled out the survey, 42.7% of those obtaining high-risk scores filled out the survey (these proportions do not differ significantly).

Also, a significant positive and strong association was found between SCC *prior* and *after* the SA ($\chi^2 (9) = 286, p < 0.001, \gamma = 0.884$). In other words, this confirms the importance of controlling for initial SCC, while analysing correlates of SCC. For more details on SCC related to background variables, we refer to Table 5.6.

In line with *hypothesis 1*, obtaining 'high-risk' scores shows to be a negative, albeit not very strong, predictor of SCC (while controlling for initial SCC) ($\beta = -0.265, t(1) = -2.36, p = 0.019$). Obtaining 'high-risk' scores⁶ explains 1% of the variance in SCC. Contrary to *hypothesis 2*, the relationship between 'high-risk' scores and SCC does not appear to depend on the initial level of SCC ($\beta = 0.05, t(1) = 0.423, p = 0.672$), meaning

⁶ The proportion of 'high-risk' scores (for those who obtained such scores) did not appear to matter ($\beta = -0.03, t(1) = -0.22, p = 0.826$).

that a high-risk score also affects those who were relatively certain about enrolling already before the SA. Contrary to *hypothesis 4*, the impact of obtaining 'high-risk' scores on SCC appeared to be moderated by gender in the sense that obtaining a high-risk score predicts lower SCC only for women ($\beta = -0.71$, $t(1) = -3.17$, $p = 0.002$). This interaction explains 1.5% of the variance in SCC. What this means in terms of the principle of fairness will be discussed in the Conclusion and Discussion section.

Enrolment. About one in three SA users (30.2%) enrolled within three months following the SA. Most of those enrolling within this period (56.5%), tended to do so within 2 days after taking the SA. This suggests that the SA was used as a 'final check'.

Obtaining 'high-risk' scores was analysed as a predictor of enrolment in a binary regression model. In contrast to *hypothesis 1*, though enrolment probability is lower for those obtaining 'high-risk' scores, compared to those who do not, the difference is not significant ($Z = -1.71$, $p = 0.086$). In regard to the mediation *hypothesis (3)*, the predictive value of SCC on enrolment (while controlling for initial SCC) turns out to be significant ($\chi^2(3) = 10.85$, $p = 0.012$, $R^2 = 14.3\%$). We compared enrolment probabilities – based on the regression model – for distinct levels of SCC by means of a non-parametric ANOVA, as the probabilities were not normally distributed. All pairwise comparisons appeared significant in which the higher the SCC, the higher the enrolment probability.

In sum, there does not appear to be a direct link between 'high-risk' scores and enrolment, yet there is a significant link between 'high-risk' scores and SCC, which in turn is related to enrolment. This indicates that there is an indirect-only mediation (Zhao et al., 2010). As described, the first path in this indirect relation ('high-risk' scores to SCC) is moderated by gender. Figure 5.3. provides a visualization of the resulting model.

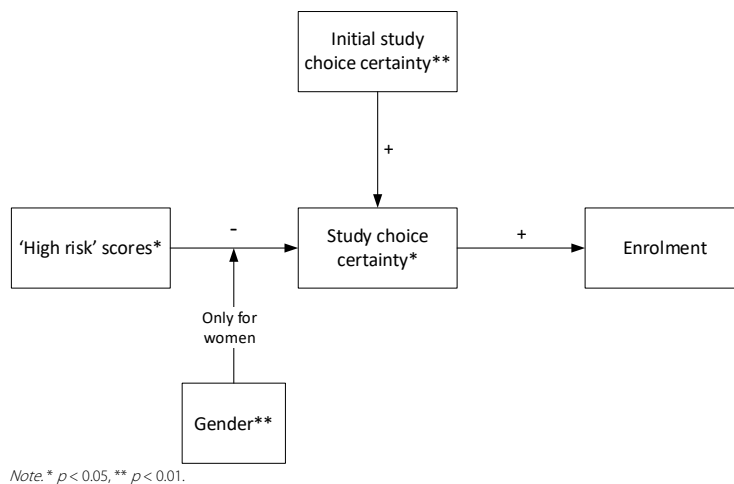


Figure 5.3. Resulting model 'high-risk' scores, SCC, and enrolment

5.3.4. Subtest scores, SCC, and enrolment

According to the hypotheses formulated in section 5.1.3., on a subtest level, we expect that subtest scores positively relate to study choice certainty (for those (rather) uncertain initially) and enrolment, independent from background variables.

Study Choice Certainty (SCC). We first included all subtest predictors and removed the non-significant subtest predictors step by step in a backwards method (beginning with the least significant). The resulting model includes the discipline test score ($\beta = 0.11$, $t(1) = 2.49$, $p = 0.014$), the numerical skills test score ($\beta = 0.12$, $t(1) = 2.84$, $p = 0.005$), as well as whether or not the expectations test is taken ($\beta = 0.28$, $t(1) = 2.61$, $p = 0.010$), while controlling for initial SCC, hours planned to study score, age, gender, and prior level of education (see Appendix B). In line with *hypothesis 1*, the significant subtest scores are positive predictors of SCC (indicated by the positive β s). Together, they explain 5.1% of the variance in SCC. In line with *hypothesis 2*, the positive effect of taking the expectations test on SCC turned out to be moderated by initial SCC ($\beta = -0.21$, $t(1) = -2.17$, $p = 0.031$), in which only those initially (rather) uncertain were affected by taking this test. This interaction effect explains an additional 0.9% of the variance in SCC.

Enrolment. Finally, the predictive value of specific subtests on enrolment was investigated in a binary regression analysis, again following a backward stepwise approach. The resulting model includes the score on the numerical skills test (Odds ratio = 1.82, $p = 0.010$) and hours planned to study score (Odds ratio = 2.02, $p = 0.005$), while controlling for discipline score, age, gender, and prior level of education (see Appendix B). In line with *hypothesis 1*, the odds ratios indicate a positive relationship between the significant subtest scores and enrolment. The resulting model explains 9.0% (Nagelkerke R^2) of the variance in enrolment.

Next, we added SCC to check whether the effects of subtests on enrolment are mediated by SCC (*hypothesis 3*). As expected, SCC itself is a positive predictor of enrolment (Odds ratio = 2.34, $p < 0.001$): the higher the certainty about enrolling after the SA, the higher the enrolment probability. Furthermore, the predictive value of numerical skills score on enrolment disappeared, once we added SCC to the equation, indicating that SCC mediates the relationship between numerical skills and enrolment. Hours planned to study score remained significant, after adding SCC (Odds ratio = 1.95, $p = 0.028$). This is not unexpected, as hours planned to study did not appear to significantly relate to SCC. In sum, there are three effects in the model on subtest level in predicting enrolment (Zhao et al., 2010):

- An indirect-only mediation of discipline and taking the expectations test on SCC and from SCC to enrolment.
- A complementary mediation from SCC on the relation between numerical skills and enrolment.
- A direct-only non-mediation for hours planned to study on enrolment.

The resulting model (see Figure 5.4.) explains 17.6% of the variance in enrolment.

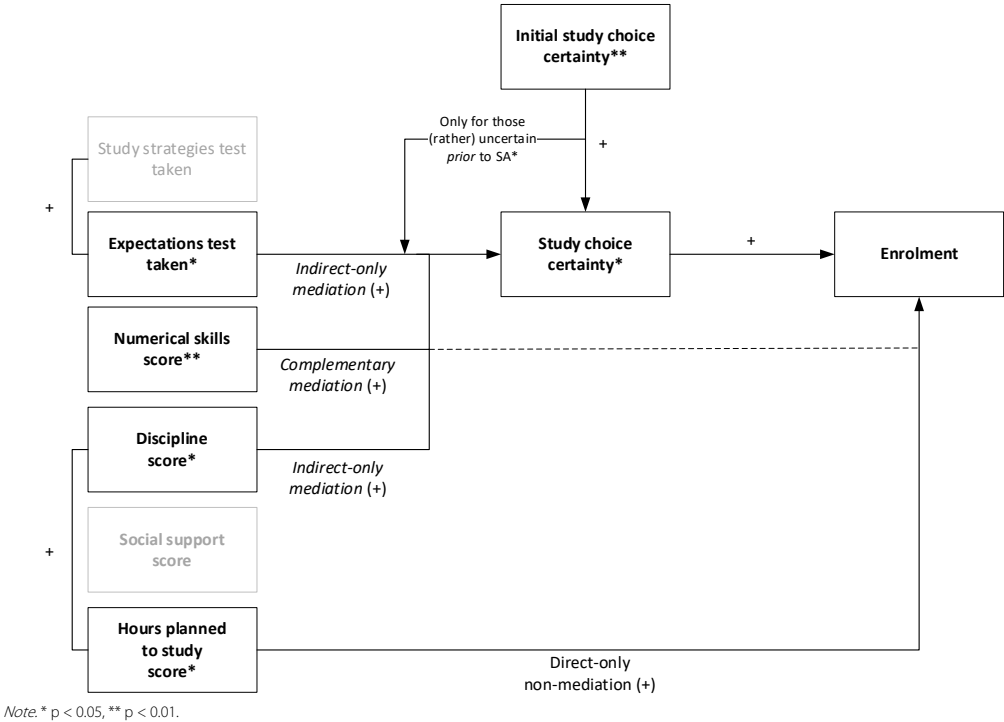


Figure 5.4. Resulting model subtests, SCC, and enrolment

Table 5.5. Subgroup division of subtests taken and % high-risk scores

Group	N	%	Subtest: % taken [% high-risk score] ¹					Study strategies	Social support	Hours planned to study	M(sd) # tests taken ²	% 'high-risk' scores ³
Men	148	41.3	100 [3.4]	17.6	91.9 [6.1]	18.2	93.1 [10.1]	93.2 [4.1]	4.14 (0.96)	6.1%		
Women	210	58.7	98.6 [3.3]	15.2	84.3 [2.4]	16.7	91.9 [7.6]	90.5 [3.3]	3.97 (1.03)	4.5%		
Elementary school	9	2.5	88.9 [0.0]	11.1	88.9 [22.2]	11.1	100.0 [22.2]	100.0 [0.0]	4.00 (0.93)	13.5%		
Pre-vocational	19	5.2	100.0 [0.0]	26.3	84.2 [15.8]	15.8	89.5 [15.8]	84.2 [5.3]	4.00 (1.29)	9.5%		
Pre-university	71	19.5	100.0 [7.0]	11.3	91.5 [2.8]	14.1	91.5 [5.6]	91.5 [1.4]	4.00 (0.97)	4.3%		
Vocational	64	17.6	100.0 [3.1]	20.3	84.4 [7.8]	20.3	95.3 [14.1]	92.2 [0.0]	4.13 (1.01)	7.1%		
Higher ed. applied	459	43.7	99.4 [1.9]	16.4	86.8 [1.3]	18.2	93.7 [3.1]	93.7 [5.7]	4.08 (0.98)	3.2%		
Higher ed. scientific	42	11.5	97.6 [4.8]	14.3	85.7 [0.0]	16.7	85.7 [19.0]	85.7 [4.8]	3.88 (1.03)	7.3%		

Note.

¹ For expectations and study strategies subtests, % high-risk score cannot be reported, as these tests do not result in a personalised score. For testing subgroup differences on subtests, we refer to Appendix B.
² Based on the total of six subtests.

³ The nominator is the number of obtained 'high-risk' scores, the denominator is the total of tests taken, based on the total four tests that result in a personalised score. For testing subgroup differences, we refer to Appendix B.

Table 5.6. Subgroup division of study choice certainty and enrolment

Group	<i>n</i> ¹	<i>SCC</i> ² <i>prior</i> to SA %			<i>SCC</i> ³ <i>after</i> SA %			% Enrolment ⁴	
		Completely uncertain	Rather uncertain	Rather certain	Completely certain	Completely uncertain	Rather uncertain	Rather certain	Completely certain
Men	103	9.7	29.1	36.9	24.3	3.9	22.3	43.7	30.1
Women	125	12.0	32.8	36.8	18.4	8.0	28.0	45.6	18.4
Elementary	6	0.0	33.3	33.3	33.3	0.0	16.7	50.0	33.3
Pre-vocational	12	8.3	16.7	58.3	16.7	16.7	8.3	33.3	41.7
Pre-university	48	12.2	32.7	30.6	24.5	6.3	29.2	37.5	27.1
Vocational	41	9.8	29.3	43.9	17.1	4.9	26.9	51.2	17.1
Higher ed. applied	104	11.5	27.9	37.5	23.1	5.7	20.0	48.6	25.7
Higher ed. scientific	20	10.0	55.0	25.0	10.0	5.0	50.0	35.0	10.0

Note:

¹ n is based on those filling out the evaluation survey, as background variables and evaluation were separate parts in the survey, not all respondents filling out background variables (see Table 5.5.), filled out the evaluation survey as well.

^{2,3,4} Differences between subgroups not statistically significant (tested in Chi-Square analyses).

5.4. Conclusions and discussion

The present study provides a hands-on example of how to investigate the consequential aspect of validity of a self-assessment for informed study decisions in (online) higher education. We investigated how the self-assessment affects prospective students' study choice certainty and subsequent enrolment. We looked into the impact of the self-assessment as a whole (obtaining 'high-risk' scores or not) as well as its constituent tests.

In regard to the *first hypothesis*, on the level of the self-assessment overall, obtaining 'high-risk' scores negatively relates to study choice certainty and, thereby (indirectly) affects enrolment (in line with *hypothesis 3*). In contrast to what was expected in *hypothesis 2*, this impact does not appear to depend on initial study choice certainty. Previous research suggested that those already quite certain about their study decision would only be looking for confirmation of their beliefs (Soppe et al., 2019), and would not be inclined to change their beliefs or decision in case of disconfirmation (Eil & Rao, 2011). In a previous qualitative (pilot) study prospective students also indicated that the self-assessment would have more impact if they would have been less certain about their decision (Delnoij et al., in review). A more diverse group of participants took part in the present study, as demonstrated by the distribution of initial study choice certainty. The results of the present study are in line with those by Van Klaveren et al. (2019), who found that the effect of risk (versus success) scores, presented in a study decision activity, was independent of being (overly) confident. In this regard, Zafar (2011) discussed that students might better be classified as Bayesian learners, who actually revise their expectations and beliefs in expected ways.

Looking at specific subtest scores, three tests (discipline, numerical skills and expectations tests) appeared to be positively related to study choice certainty (*hypothesis 1*). The relation between the expectations test and study choice certainty exists only for those initially (rather) uncertain about enrolling, which is in line with *hypothesis 2*. Though, in contrast to the hypothesis, this is the only effect was moderated by initial study choice certainty. Nevertheless, this still indicates that even without a personalised score, tests might be of relevant added value, in this particular case helping those who are initially not very certain yet about their study decision. Furthermore, on the subtest level, two tests (hours planned to study and numerical skills) related to enrolment, in which the relationship between numerical skills score and enrolment is mediated by study choice certainty (in line with *hypothesis 3*). The social support test did not turn out to be significant in relation to study choice certainty and/or enrolment. More than half (54%) of the self-assessment users in this study took this test, and it also appeared the test that most often (compared to the other tests) results in a 'high-risk' score. Yet, that does not seem to affect the study choice (certainty) of prospective students. One of our previous (pilot) user studies showed that prospective students rate the relevance of this test for their study decision relatively low (Delnoij et al., 2020b). Perhaps this result generalizes to the more diverse group of participants in the present study, explaining the lack of impact. Literature does suggest social support to be a relevant factor for study success, especially in the context of distance (*online*) higher education (Delnoij et al., 2020a; Muljana & Luo, 2019). This implies that due attention should be paid to clarifying

this relevance for prospective students to effectively engage with this test and feedback in the self-assessment.

In regard to fairness, there are two specific points to discuss. First, of those not obtaining any high-risk scores, 4.1% (8 participants) still became less certain about enrolling, implying a threat to consequential validity. As they obtained significantly more 'medium-risk' scores than those not obtaining any 'high-risk' scores and not discouraged (see section 5.3.3.), we might conclude that this does not actually indicate much of a threat to consequential validity. Furthermore, the expected (negative) relation between obtaining 'high-risk' scores and study choice certainty appears to exist only for women, while compared to men, they did not obtain more 'high-risk' scores in general or on specific subtests. This implies a threat to consequential validity. A possible explanation for this result might be found in gender differences reported in attribution theory (Beyer, 1998). This theory suggests that men tend to attribute poor performance (e.g., a 'high-risk' score) to lack of effort, whereas women are more likely to 'doubt themselves'; to attribute poor performance to a lack of ability. Though this effect explains only 1.5% of the variance in study choice certainty which begs the question whether and how this degree of 'unfairness' should be addressed. In this respect, it is important to note that in fact high-risk scores were meant to prompt prospective students to carefully reflect on their study decision. In this respect, the effect found for women is not 'unjustified' or unfair, but the lacking of this effect for men is. This implies that further research is needed to investigate how the feedback provided to men should be adapted to ensure that they do not dismiss high-risk scores too easily.

All in all, this study shows that a self-assessment for informed study decisions does affect study decisions in terms of study choice certainty and enrolment behaviour. In general, the demonstrated impact is in line with the purposes of the self-assessment and therefore, constitutes evidence for the consequential aspect of validity. A small (in effect size) threat to the consequential aspect of validity was demonstrated by the insight that men do not appear to be influenced by obtaining 'high-risk' scores, whereas women do. In general, the effects (e.g., in terms of explained variance) of the SA on study choice certainty and enrolment are relatively small. However, the added value of the SA should not only be derived from this, but should also be viewed in light of cost-effectiveness and scalability (Kraft, 2020). In this regard, it is also important to note that the SA does seem to have added value for a reasonably large group (68.9%), in line with purposes of the test. This is expressed, for example in the fact that prospective students tend to become more certain about enrolling if they do not obtain any 'high-risk' scores and – to a certain extent – seem to adequately downsize their certainty in case they do obtain such scores.

The present study modelled how to investigate the consequential aspect of validity. In general, this aspect of validity has been underexplored in applied validation studies (Cook et al., 2014). Especially, in the context of study decision tools (Niessen & Meijer, 2017) and even more so in educational contexts with broader accessibility, like *online* higher education (Soppe et al., 2019).

5.4.1. Limitations and recommendations for future research

The predominant limitation is that not all subtests in the self-assessment evaluated in the present study were fully developed. For two tests (expectations and study strategies), a score distribution was not yet set, which means prospective students did not obtain personalised feedback on these tests. Those who took the study strategies test also took the expectations test, which means we could not investigate both tests separately, as this would result in multi-collinearity. We included the test completed most often – the expectations test – as a predictor, but exactly which part of the effects in regard to that predictor can be attributed to that test and which to the study strategies test needs further investigation. Therefore, and in line with the iterative design-based validation approach adopted for the present self-assessment, a more thorough consequence evaluation of these tests is needed, once a score distribution has been determined and feedback provided accordingly. After all, validation is not a once and for all call, but involves ongoing monitoring in light of possible changes occurring in the context and fluctuations in the target population. Replication of the current results as well as more longitudinal consequence evaluation would therefore be valuable. The latter especially, to capture the consequential aspect of validity more fully, by complementing the current findings with data regarding students' success after enrolment, in order to re-establish the adequacy and predictive power of the tests already fully developed and develop the other tests to fully personalised versions. Though, we must bear in mind the more general limitation that we can only evaluate consequences for those who did decide to enrol, be it on the base of favourable assessment results or despite unfavourable results. For those who do not enrol, we will never know whether they would have been successful. If it was an unfavourable test result that led them to decide not to enrol, we will never be able to tell whether the assessment unjustly led them to re-think and decide otherwise or safeguarded them from a frustrating and possibly painful experience.



General discussion



The research in this thesis involved the design and evaluation of an online self-assessment for informed study decisions in online higher education. The self-assessment aimed at stimulating reflection and providing feedback for action (e.g., remediation) to support prospective students in making well-informed study decisions. The ultimate goal being to decrease the non-completion problem in this context, by creating resources for retention in an early (pre-enrolment) stage.

The main objective of this thesis was to evaluate five sources of validity evidence in regard to the self-assessment: content, predictive, internal structure, process, and consequential aspects of validity. Thereby, we addressed the lack of empirical foundation of self-assessments aimed at informed decision making in higher education. So far, such instruments were rarely transparently designed or validated (Niessen & Meijer, 2017) and the limited number of validation studies focused merely on internal structure and predictive aspects of validity. The present thesis advocated for evaluating all sources of validity evidence and modelled how to do so in a design-based validation approach, linking the five validity aspects to common phases of design-based research: analysis, design and development, and evaluation.

This general discussion shortly summarises the results of the present research, discusses the implications in light of design-based research contributions, and is concluded by an agenda for future research, based on reflections on the limitations of the present research.

Summary

The thesis reports on the results and approach of developing the self-assessment in three parts, according to design-based research phases.

Part I – Analysis (Chapters 1 and 2) focused on analysing the non-completion problem in (online) higher education. Chapter 1 involved a double systematic review resulting in an overview of predictors of completion and characteristics of interventions that address non-completion. The conclusion that pre-enrolment interventions did not received as much scientific attention in combination with the fact that such interventions seemed a promising approach to tackle the non-completion problem (e.g., see Muljana & Luo, 2019) led to the decision to develop a self-assessment for informed study decisions. The predictors of completion as revealed in the literature review constituted the first steps in determining the content of the self-assessment. These predictors were further investigated in terms of predictive value and internal structure in the specific context of *online* higher education in a follow-up correlational study (Chapter 2). Based on these two chapters, a first step prototypical self-assessment was developed.

Having established an evidence-informed prototype of the self-assessment, **Part II – Design & development (Chapter 3 and 4)** focused on small-scale user tests to evaluate potential users' expectations of and experiences with the self-assessment. This part mainly focused on process and consequential aspects of validity. Insights on how prospective students proceed through and react to the self-assessment were indispensable and resulted in major adaptations of the self-assessment. For example, the

design of the feedback – consisting of information and a visualization of the obtained score, an explanation of the measurement and an advice for further preparation – was based on the results of these user studies. Also, these studies provided insight from the user perspective on the content of the self-assessment as well, leading to an expansion of subtests.

Part III – Evaluation (Chapter 5) provided summative evaluative insights, based on data collected during actual full-scale deployment of the self-assessment. In this evaluation, prospective students took the self-assessment in an authentic situation of orienting towards studying at the Open University of the Netherlands (OUNL). This evaluation focused on the relationship between the self-assessment, study choice certainty, and enrolment. Thereby, this part shed further light on the consequential aspect of validity. The conclusion of this large-scale evaluation was that the self-assessment appears to affect study choice certainty, and, thereby, enrolment probability for a relatively large group of prospective students and in line with its purposes. Based on these results, recommendations for further development and implementation of the self-assessment as well as for future research were made.

Contributions and implications

Whereas the thesis chapters and summary provide descriptions of the main findings in more concrete terms, the next sections discuss those findings in light of design-based research contributions. Although design-based research is directed at finding solutions for complex problems in specific contexts, it also aims to amplify the more general body of knowledge, i.e., to contribute to theories in the field (Edelson, 2002). Edelson describes four features that distinguish design research from simple design and that augment useful lessons to be applied beyond the specific context of the design: research driven, systematic documentation, formative evaluation, and generalization. This thesis is a demonstration of all four, but this concluding chapter focuses on the generalization component, retrospectively – overseeing the whole development process of the self-assessment:

“It is through the process of generalization that a design researcher takes the specific lessons of one or more design experiences and contributes to the development of domain theories, design frameworks and design methodologies” (Edelson, 2002, p. 117).

The implications of the present research are discussed in light of these three kinds of contributions – domain theories, design frameworks and design methodologies.

Domain theories are descriptive in nature, as they focus on real-world issues and processes, not on design per se. Domain theories can be further characterized as outcome theories or context theories. *Outcome theories* describe the (desired) outcomes associated with a design / an intervention. Understanding the desired outcomes of implementing an intervention is essential to the successful design of this intervention (Edelson, 2002). In the present thesis, the ultimate goal is to enhance completion rates in higher education. It is therefore important to understand the non-completion problem first and to target the self-assessment at relevant factors related to completion.

Complementary to domain theories, *context theories* describe challenges and opportunities pertaining to the context in which a design (intervention) is to be implemented. In light of the present thesis, context theories shed light on the self-assessment in the context of the study decision-making process as a whole.

The second set of contributions contains ***design frameworks***, which are prescriptive and indicate requirements or characteristics of a particular intervention for a particular purpose. An example of a design framework contribution would be design guidelines for learning analytic dashboards to enhance self-regulated learning as described in the thesis by Jivet (2021). Based on the results of the present thesis, guidelines are formulated for the content, score distributions and providing feedback in self-assessments for informed study decisions.

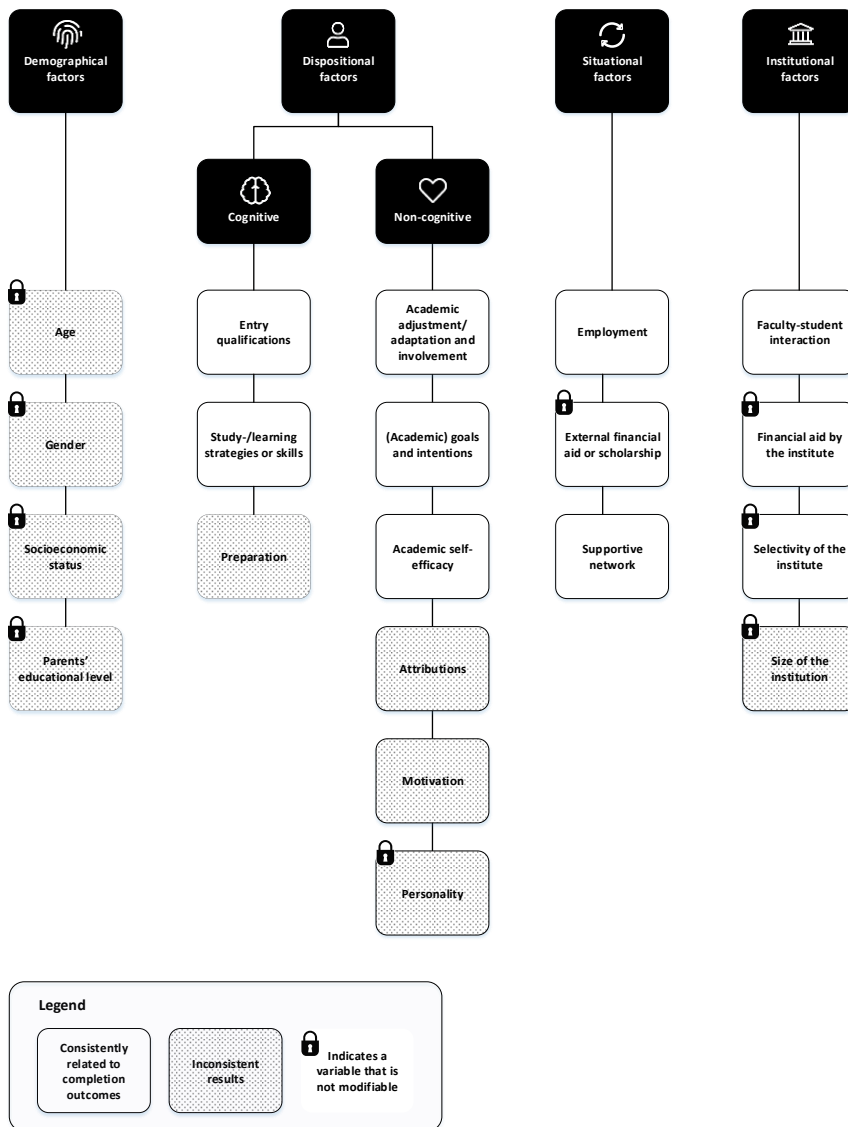
The last category of contribution entails ***design methodologies***. Design methodologies are also prescriptive in nature, yet not focused on the design itself, but on the procedure in order to arrive at a design. A design methodology typically describes the recommended processes and stakeholders to be involved in order to reach a certain design. As Edelson describes, an example of such a contribution can be found in user-interface design in computer science, in which numerous design methodologies have been created to make sure that data and feedback from users are obtained at appropriate intervals and incorporated into design. In the present thesis, this involves the integration of design-based research stages with the five sources of validity evidence, derived from modern validity theories.

The next sections reflect on the implications of the present thesis in light of these types of contributions by answering the overarching question: *what implications does the work described in this thesis hold for theories of this specific type?*

Domain theories – *Outcome theories*

The ultimately desired outcome related to the design-based research described in this thesis is completion in higher (online) education. In that regard, the first chapter in this thesis builds on a vast body and broad variety of prior research. For years, predictors of completion have been studied in isolation (e.g., Pinxten et al., 2019; Van Herpen et al., 2017), combined in review studies (e.g., Robbins et al., 2004), and summarized in theoretical models (e.g., Tinto, 1975; Neuville et al., 2007). Because such a wide range of research was already available, our focus was on creating an overview thereof. The quite basic classification of Carroll et al. (2009) proved a useful vehicle for creating the overview; distinguishing predictors as *dispositional* (i.e., individual factors, internal to the student), *situational* (i.e., related to the circumstances in students' particular lives), or *institutional* (i.e., factors resulting from procedures, policies and structures of an educational institution). Based on our findings, we proposed to make a further distinction within the category dispositional factors between *cognitive* (i.e., ability, knowledge, skills and experiences) or *non-cognitive* (i.e., affective and attitudinal factors). This distinction is particularly relevant with an eye on developing interventions to tackle the non-completion problem, as they possibly require different approaches. Also with an eye on

developing interventions, Chapter 1 added to existing models by introducing predictive consistency and modifiability as relevant characteristics of predictors and integrating them in the resulting model (see Figure 6.1). A variable is considered modifiable when it is changeable and/or can be advised upon. For example, study skills are trainable (Patterson et al., 2014) and the intended division of hours to spend on employment, study, and other activities can be advised upon (though not always changed).



Note. This Figure is designed in Microsoft Visio®

Figure 6.1. Predictors of completion in higher (online) education

Domain theories – *Context theories*

For the present research, the context is online self-assessments as part of the study decision-making process in higher online education. Several theories are relevant here, with varied foci on either online self-assessment or the study decision-making process. Prior to connecting findings of the current study to related theories, several observations must be made.

Firstly, investigations of the study decision-making process in *online* higher education are relatively scarce compared to higher education more generally. Though self-assessments for informed study decisions are deployed in both settings, the distinction between these settings is not simply a matter of different modes of delivery, but more importantly, of different student populations. The study decision-making process of adolescents in initial education (i.e., prior to entering the labour market) can be expected to substantially differ from that of adult learners. To illustrate this, for adolescents the question whether or not they will enrol is not so much an issue. All in all, this means that the current study's contributions to existing context theories are likely to be limited and to take the form of 'tentative comparative reflections'.

Secondly, a self-assessment for informed study decisions constitutes 'just a step' in a range of activities prospective students may (or may not) carry out in the entire decision making process, which indeed, may be further conceptualised as part of career development or career decision making (Germeijs & Verschueren, 2007). With this broader scope comes a more prominent role for matching personal interests with subject choices, especially in the case of adolescent prospective students. However, Lent and Brown (2013) warn against the focus of most career development theories on the content questions ("the destination rather than the journey" p. 557). The Social Cognitive Model of Career Self-Management the authors propose is meant to support a shift of focus to process aspects of career behaviour, independent of specific educational and occupational fields. In this regard, Chapter 2 has explicitly focussed on generic predictors of completion, independent of subject choice.

Despite the more limited scope of the current study, some (dis)similarities are noteworthy. First, this dissertation revealed an important challenge for (online) self-assessments for informed study decisions. This challenge involves the timing of the self-assessment as an orientation activity in the study decision process. Chapter 4 demonstrated that if prospective students take the self-assessment when they are already reasonably certain of their decision, they perceive the self-assessment as less relevant. These students indicated to look mainly for confirmation and appeared somewhat resistant to reconsider the choice they 'already made' (Chapter 4). A similar result was found by Germeijs and Verschueren (2007), who investigated career decision making including study decisions in students leaving high school until their second year in higher education. Related to our concept of study choice certainty, they examined commitment to the study decision and found that initial commitment to the study decision explains most of the variance in commitment towards the end of the study decision process. Although, it can be argued that confirming a choice that was already reasonably certain would still be a 'valid' contribution of the self-assessment to study decision making

(Soppe et al., 2019; Chapter 4), Chapter 5 revealed that confirmation is not the only contribution of the self-assessment for those initially certain. In this study, we expected that those students who were (rather) certain about enrolling were more likely to remain unperturbed by less favourable self-assessment results, than those initially (rather) uncertain about enrolling. This did not appear to be the case. Results indicated that the self-assessment could still have an effect even when people are already quite certain. Of those initially certain and obtaining 'high-risk' scores 30% adequately downsized their certainty. In this regard, Zafar (2011) discussed that students might better be classified as Bayesian learners, who actually revise their expectations and beliefs in expected ways. Still, timing of the self-assessment in the decision making process is important, but – at least in the case of open distance education – ultimately controlled by prospective students themselves.

Another important addition of the present research is the expectations subtest. Although we have not yet been able to validate this test in terms of internal structure and predictive value, Chapter 5 showed the relevance of this subtest, especially for those prospective students who were initially not certain of their study decision yet. In other words, our study showed that the expectations test is conducive to study choice certainty and, thereby, to enrolment probability. We did not investigate the relationship with post-enrolment success, but McGhie's (2017) in-depth longitudinal study suggests that successful students differ from less successful students in terms of realistic expectations. Although McGhie's study took place in a different context (first year in South-African higher education), the study suggests that holding clear and realistic expectations is conducive to success after enrolment.

Design frameworks

Based on the research in this dissertation, this section addresses guidelines in regard to the content of the self-assessment, score distributions, and the feedback.

Self-assessment content. Chapter 1 demonstrated that scientific attention was lacking for interventions aimed at enhancing completion in higher (online) education, taking place *prior* to student enrolment. The present dissertation therefore focused on translating predictors of completion into a pre-enrolment intervention aimed at enhancing completion: a self-assessment for informed study decisions. For such self-assessments to effectively contribute to enhancing completion, they should be targeted at *modifiable* and *pre-enrolment relevant predictors* of completion. Modifiability is required as the self-assessment aims (a) to inform prospective students on possible room for improvement in regard to their study preparedness and (b) to provide feedback for action so prospective students can further prepare and enhance their chances for completion. Also, the subtests in the self-assessment should target factors that are relevant prior to enrolment. After all, it would not make sense for pre-enrolment interventions to target variables that, in nature, can only play a role after enrolment (e.g., academic adjustment).

Moreover, the relevance of subtests should also be clear to those using the self-assessment. Though research and theory might imply certain tests (factors) to be relevant (predictive of completion), this does not necessarily mean that users perceive this

relevance too. If prospective students do not see this relevance, the likelihood that they engage with the tests and aligned feedback decreases (King & He, 2006; O'Donovan & Smyth, 2005). In the present self-assessment, relevance of subtests is explicitly addressed in feedback. However, this does not withhold prospective students from ignoring certain subtests in case they might not seem as relevant at first sight. So, the challenge is to clearly communicate relevance beforehand without inciting/steering strategic test-taking behaviour (Viswaveran & Ones, 1999). Of course, the risk of strategic test-taking behaviour is less of an issue in low-stakes assessments (Sjöberg, 2015).

Score distributions. A score distribution provides context and meaning to possible scores on a test and indicates the cut-off point below which remediation or further preparation is considered beneficial (Cook et al., 2014). Such a score distribution can be set in various ways, but should be well considered and aligned to the specific context in which an assessment is implemented (Downing, 2003). Nolden et al. (2019), for instance, created three categories of scores based on the means and standard deviations of (non-)successful students. A 'high-risk' score in their distribution entails all scores lower than the mean of non-successful students (in a previous cohort) minus one standard deviation. A 'low-risk' score entails all scores higher than one standard deviation above the successful students' mean score. 'Medium-risk' scores are all scores in between these two 'extremes'. We chose to base our score distribution on a classification model with rather strict sensitivity (Chapter 2), to minimise the likelihood of false negatives. More specifically, this means that a maximum of 5% of students receiving a '*high-risk*' score on a certain test might in fact appear completers. This approach enabled us to take into account the accessibility of open online education, meaning that we wanted to be particularly careful not to unnecessarily discourage students. When applying Nolden's et al. strategy, we would end up with relatively more false negatives, not in line with the open access context of our institution. So, if minimizing false negatives is desired, a strict sensitivity should be maintained.

Feedback. There is a broad variety of literature available on feedback, mainly in relation to learning. The principles for good feedback proposed in that context focus on improving the learning process and learning outcomes (e.g., Hattie & Timperley, 2007). Feedback objectives in this context do not completely fit with the objectives of the present self-assessment. The purpose here is not to enhance learning, but to inform decision-making. In this regard, the first aim of the feedback aligned to the self-assessment is to provide *food for thought* or, in other words, to elicit reflection on study preparedness. In Chapter 4, it was demonstrated that the self-assessment meets this aim as prospective students reflected on their preparedness and how they might improve this. Moreover, the feedback made prospective students aware that they could influence their skills and preparedness. This resulted in enhanced levels of self-efficacy – a person's sense of their own ability to accomplish something successfully (Bandura, 1977). Most importantly, such self-efficacy is an important determinant of students' motivation (Ryan & Deci, 2000) and success in higher online education (Harnett, 2016).

The second aim of the self-assessment was to provide *feedback for action*. For feedback to be accepted and used, the process of reflection appeared instrumental in

previous findings (Sargeant et al., 2009). Though, awareness of issues does not necessarily imply that beneficial follow-up actions are being taken (Jivet et al., 2017). In this regard, Chapter 4 demonstrated that the intention to take follow-up actions depended on self-assessment scores and initial study choice certainty. Though participants in this study reflected on their preparedness, most rated their intention to actually take follow-up actions rather low. They explained that this was mainly because they scored relatively well and, therefore, did not feel the need to take follow-up actions, and because they were already quite certain about their decision before taking the self-assessment. In line with research by Soppe et al. (2019), those initially certain appeared to be looking mainly for confirmation.

Finally, in regard to the feedback, a specific design decision in the development of the feedback for the self-assessment concerned whether or not to use social comparison in presenting the obtained scores. Jivet (2021) warns that such motivational triggers might determine prospective students' definition of success and the way they approach their goals. These triggers could lead to adopting a performance orientation to goals in which students either focus on doing better than others or avoid doing worse (Elliot & Harackiewicz, 1996). In contrast, a mastery approach to goals, means students focus on developing knowledge and skills, without a relation to others (Elliot & McGregor, 2001). Such mastery-oriented goals appear more favourable for learning and achievement; whereas performance oriented goals in some cases even seem to have an adverse effect (Huang, 2012). In sum, design decisions might implicitly communicate underlying values to prospective students. Designers need to be aware of their responsibility in designing such tools for (prospective) students and approach this with caution and empathy (Jivet, 2021). In this regard, formative evaluation of design decisions involving user tests are very important (Edelson, 2002). In such a (small-scale) user study (Chapter 3), we showed that prospective students' in the context of our self-assessment did not necessarily value a comparison of their obtained score to the score of a reference group. The present self-assessment focuses on adult students, who appear less in need for social comparison with peers (Callan et al., 2015; Festinger, 1954).

Design methodologies

As explained before, we integrated a design-based research methodology with state-of-the-art validation theory for developing self-assessments for informed study decisions in higher (online) education. As visualized in Figure 6.2, five sources of validity evidence (Table 6.1) were linked to stages and characteristics of design-based research, i.e., an iterative process of analysis, design, and evaluation (Van den Akker et al., 2013).

The **analysis stage** focused on content, internal structure and predictive aspects of validity by shedding light on what factors are related to completion and, thus, should be targeted in the self-assessment. As mentioned, the present dissertation adds to prior research on predicting completion by focusing on predictive consistency, modifiability and pre-enrolment relevance of factors (Chapter 1 and 2). This stage further indicated how such factors could be measured in a self-assessment to detect students at risk for non-completion. An important note here is that existing tests and questionnaires cannot simply be copied from one context to another (AERA et al., 2014; Royal, 2017). In line with

this, the research in this dissertation (Chapter 2) showed that the internal structure of instruments depends on the specific context for which they are used. For the internal structure validity aspect, as well as for the predictive aspect, there are applied validation studies focused on similar tools for informed study decisions in other contexts, e.g., see Germeijs and Verschueren (2006), Lui et al. (2018), and Nolden et al. (2019). Looking at other validity aspects, specifically the process (i.e., how users proceed through and react on the assessment) and consequential (i.e., impact on desired outcomes) aspects of validity, it turned out that explicit attention is lacking in the development of self-assessment for informed study decisions.

Therefore, in the ***design and development stage***, the focus shifted to the process and consequential aspects of validity. In addition, insights on the content aspect of validity were supplemented from a user perspective. The process and consequential aspects of validity were addressed, by providing insight in how prospective students proceed through and react on the self-assessment and gain insight in the extent to which the self-assessment affects study choice certainty and intentions for further preparation. These evaluations were crucial for the continued development of the self-assessment. Interestingly, the research in this dissertation revealed contradictory requirements for the length of subtests. From the parsimony requirement, subtests should be as short as possible. However, some prospective students felt that with a limited number of items, their score and related feedback were not as sound, which made them doubt the accuracy of the discipline and social support tests (Chapter 4). Downing and Haladyna (2004) already discussed the concept of construct under-representation as a threat to validity. The present dissertation adds to that by making explicit the concept of *perceived* construct under-representation. Though statistically and theoretically a construct might be well represented in a certain test, this is not necessarily how users will perceive it too. It is important to take into account such experiences and perceptions as this might influence prospective students' engagement with the feedback and intention to take follow-up actions based on that feedback. As said, relatively rapid and low-cost pilot tests (Chapter 3 and 4) enabling formative evaluation of the intervention are therefore crucial in the design process (Edelson, 2002).

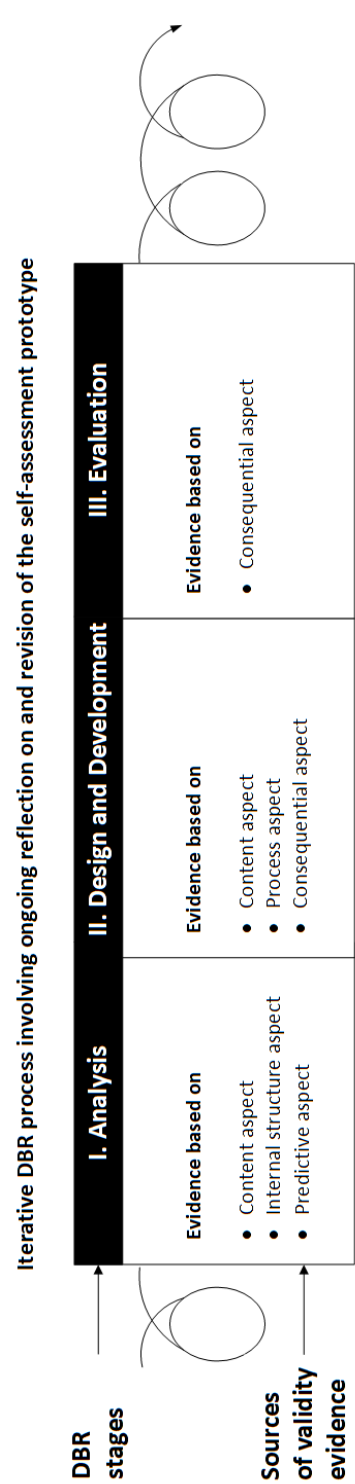


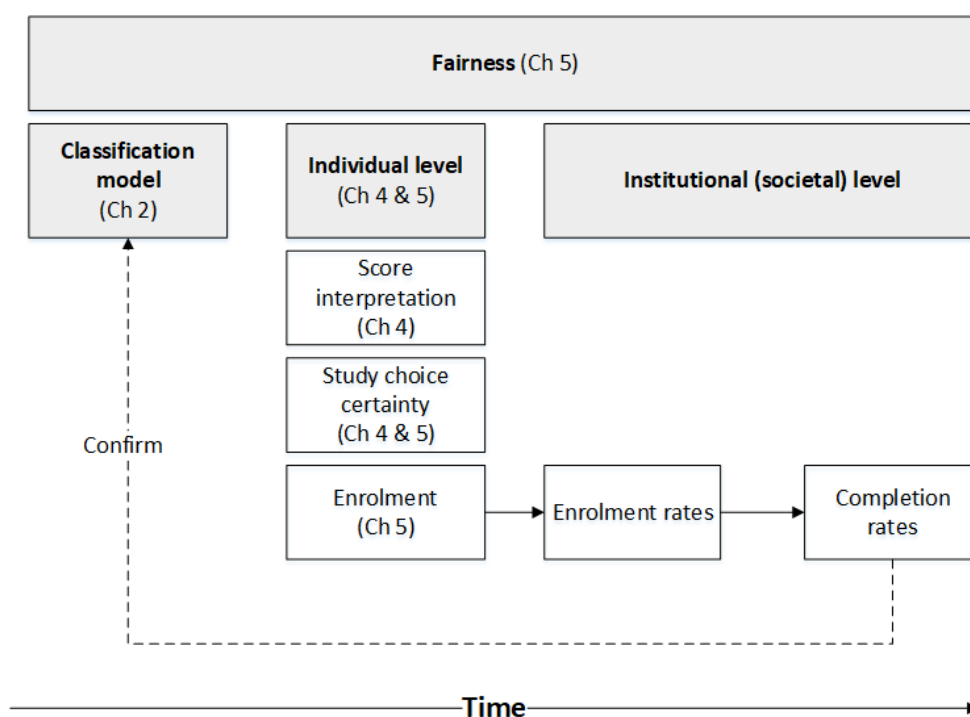
Figure 6.2. Design-based validation model to develop self-assessments for informed study decisions

Table 6.1. Aspects of validity evidence

Validity Aspect	Explanation
Content	Extent to which the test content accurately represents the content domain.
Predictive	Relative performance of test scores in predicting (supposedly) related variables.
Internal structure	The degree to which items reflect coherent dimensionality both on theoretical and statistical grounds.
Process	The fit between what the items/test intend to measure and the experiences/considerations test takers take into account when responding.
Consequential	The degree to which interpretations of and actions following test results/feedback are in line with the intended purposes.

Note: Based on AERA et al., (2014), Beckman et al. (2005) and Cook et al., (2014).

Finally, in the **evaluation stage**, the consequential aspect of validity was the core subject. The consequential aspect of validity regards the intended and unintended impact of the assessment, both positive and negative, for the individual and/or society, i.e., the soundness of decisions made and actions taken, based on assessment results (e.g., taking a remediation course to address sub-standard performance)(AERA et al., 2014; Beckman et al., 2005; Cook et al., 2014). This validity aspect spreads over several stages as visualized in Figure 6.3, in which three components of consequential validity are addressed: *pass/fail cut-off establishment and consequences*, *individual and societal impact of assessment scores*, and *fairness* (AERA et al., 2014; Beckman et al., 2015; Cook et al., 2014; Downing, 2003; Messick, 1995; Niessen, 2018; Schreurs, 2020; St-Onge et al., 2017).



Note: This Figure is designed in Microsoft Visio®

Figure 6.3. Longitudinal perspective on consequential validity

As explained in the design frameworks, the pass/fail cut-off establishment was based on the classification model with a rather strict sensitivity (Chapter 2). To fully capture the consequential aspect of validity (impact on institutional level) and to confirm the classification model as set in an earlier stage (see Figure 6.3.), longitudinal evaluation is required of students' success after enrolment. Here, we must beware that we can only evaluate the classification model for those who did decide to enrol, either following favourable assessment results or despite unfavourable results. For those who do not enrol, we will never know whether they would have been successful. If it was an

unfavourable test result that led them to decide not to enrol, we will never be able to tell whether the assessment result justly led them to decide otherwise or unjustly discouraged them. In other words, fully confirming the classification model set earlier in developing the self-assessment is thus, impossible.

To evaluate the impact on individual level, Chapter 4 models how to capture prospective students' score interpretations and intentions for follow-up orientation/preparation. As discussed, the latter seemed to depend on the level of study choice certainty, which was further investigated in Chapter 5. That study, on a larger scale, showed that the self-assessment affects study choice certainty in line with its purposes for a reasonably large group of prospective students (68.9%). Their study choice certainty was adapted or remained unchanged in accordance with their obtained scores.

Last but not least, fairness is an important, yet underexplored component of the consequential aspect of validity (Kreiter, 2016; Xi, 2010). It has been defined as "comparable validity for identifiable and relevant groups across all stages of assessment, from assessment conceptualization to the use of assessment results" (Xi, 2010, p. 154). In the present thesis, the gap in the literature was addressed by comparing the impact of the self-assessment across different subgroups, based on age, gender and prior level of education. In Chapter 5, it was concluded that the impact of the self-assessment on study choice certainty and enrolment decisions can be considered fair. One small (in effect size) threat to fairness was found by the result that the negative relation between obtaining 'high-risk' scores and study choice certainty appeared to exist only for women. Compared to men, they did not obtain more 'high-risk' scores in general or on specific subtests. In this respect, the effect found for women is not 'unjustified' or unfair, but the lacking of this effect for men is. This implies that further research is needed to investigate how the feedback provided to men should be adapted to ensure that they do not dismiss high-risk scores too easily.

As can be derived from the summary above, a chronological order appears to exist when it comes to collecting evidences for the validity aspects: investigating process and consequential aspects of validity make sense after content, internal structure and predictive aspects have been more or less secured. However, we must bear in mind that investigating validity is not a 'once and for all' activity, but one that requires regular attention, as student populations and/or educational practice evolve over time (Messick, 1989; Royal, 2017).

Limitations & directions for future research

The present thesis approaches study decision making from a process perspective, including initial study choice certainty, sub-tests taken, scores obtained, subsequent study choice certainty and enrolment (Chapter 4 and 5). However, we did so, only for the group of prospective students taking the online assessment. Comparing their study decision process with that of prospective students not taking the self-assessment was not part of the current study. Clearly, this merits future research in the continuous investigation of evidences of the consequential aspect of validity of the self-assessment. Context theories in the field of career decision making suggest that self-efficacy as well

as outcome expectations for career exploration and decision-making determine engagement in self- and career exploration actions (e.g., taking an online self-assessment for study choice decision making), as well as the outcomes of this process (decidedness) (Lent & Brown, 2013). Hence, these variables should be included as possible discriminants of self-assessment takers versus non-takers. At any rate, prospective students included in the study described in Chapter 4 of this thesis indicated they found the self-assessment made them aware that they can control their level of preparedness. These findings suggest an alternative causal relation might be hypothesized as well.

In addition, recent research has demonstrated that the study decision process does not have a clear end point and that it is an ongoing process of meaning making, which continues even after students are enrolled in higher education (Vulperhorst et al., 2021). Therefore, we recommend future research to focus on alignment of the present self-assessment to continued post-enrolment advice.

Another limitation is that the evaluation of the extent to which the self-assessment fulfils its purposes focused merely on the *food for thought* (reflection on preparedness) component of this aim. Though Chapter 5 provides insight in the *feedback for action* component to some extent, by investigating enrolment behaviour, the present thesis has not taken into account whether prospective students follow up on the advice for further preparation (i.e., consult the study advisor or make use of the links to online resources to improve certain skills). Future research is necessary to investigate these follow-up actions and the extent to which they affect success after enrolment. Though, this latter part comes with the inherent constraint that this can only be investigated for those ending up enrolling.

Finally, the ultimate goal aligned to the self-assessment is to decrease the non-completion problem in online higher education. Evaluating completion rates requires longitudinal evaluation and was beyond the scope of the present thesis. In the ongoing validation of the self-assessment, this should be one of the next step towards fully capturing the effectiveness of the self-assessment in informing study decisions and building resources for completion.

Concluding remarks

This thesis started with the aim of contributing to decreasing the non-completion problem in higher (online) education. Specifically, this led to the development and evaluation of a self-assessment for informed study decision, in order to build resources for completion in an early stage. One year after the implementation of the self-assessment, approximately 2000 prospective students have used the self-assessment (i.e., completed at least one subtest) and generally appeared to find it useful in the study decision process (i.e., 86.6% of those who filled out the evaluation survey ($n = 535$) find it useful). Besides the fact that prospective students appreciate the instrument, it is also important that the self-assessment fulfils its purposes by actually informing students in regard to their study decision and supporting them to start their studies in higher education well prepared. The five chapters in this thesis provide insight into the process of collecting and evaluating sources of validity evidence, to more or less secure that objective. Thereby, the research in this dissertation adds to the literature on

(self-)assessments in the context of orientation and admission towards higher education, as such assessments are rarely thoroughly (and publicly) validated (Niessen & Meijer, 2017). In addition, this research adds to the validity literature, by providing a hands-on example of applied validation studies for *all* validity aspects, which – so far – tend to focus mainly on high-stakes assessments (i.e., selection, pass/fail, or grade), standardized tests, and predominantly in the context of health professions (Cook et al., 2014; Wools et al., 2010).

Access to higher education – even if (or especially when) it is open – requires the best possible decision making support. Hence, it is a call of duty to justify assessment procedures in this context, based on careful design, continuous evaluation, and empirical arguments.



Appendices



Appendix A – Quality of review studies on predictors of non-completion (Chapter 1)

Reference	Databases and search terms used provided	In- and exclusion criteria provided	Number of studies included ¹	Independent variables definition provided	Independent variables operationalization provided	Outcome measure definition provided	Transparency on generalizability
1	no	no	7	no	yes	no	yes for both
2	yes	yes	109	yes	yes	yes	yes for both
3	yes	yes	7	no	no	no	yes for both
4	yes for databases only	yes	58	yes	yes	no	yes for included studies
5	yes	yes	35	some	some	no	yes for both
6	yes	yes	13	no	some	no	yes for both
7	yes	yes	273	yes	yes	yes	yes for both
8	yes for databases only	yes	34	no	no	no	yes for both
9	yes for databases only	yes	58	yes	yes	no	yes for both
10	yes	yes	39	yes	some	yes	yes for both

Note: ¹ In relation to the non-completion outcome measure; ² Of included studies and review study itself

Appendix B – Quality of intervention studies (Chapter 1)

Reference	Sample size (total)	Sampling method provided	Target factor for the intervention explicitly aligned to predictors of non-completion	Transparency on generalizability	Transparency on threat to internal validity
1	unknown	partly	yes	no	no
2	139	yes	yes	no	no
3	53	yes	no	no	no
4	236	yes	no	yes	yes
5	128	yes	yes	yes	no
6	153	yes	no	yes	yes
7	16175	yes	yes	no	yes
8	307	yes	yes	yes	yes
9	453380	yes	no	yes	yes
10	1618	yes	yes	yes	yes
11	13555	yes	yes	yes	yes
12	4019	yes	no	yes	yes
13	4681	yes	no	no	yes
14	2700	yes	yes	yes	no
15	3374	yes	no	no	yes
16	1486	yes	yes	yes	yes

Appendix C – Results on predictors of non-completion in higher education (Chapter 1)

Category	Reference	Educational context	Factor	Outcome measure formulation ¹	Results and effect(s) ²
Demographic variables	2	Traditional higher education	Socioeconomic status	Completion	Significant Direction: positive K: 6 N: 7704 Effect: estimated true correlation = .228
	6	Traditional higher education	Gender	Non-completion	Significant in K: 3/10 Non-significant in K: 7/10 Direction: diverse N: - Effect: -
	6	Traditional higher education	Age	Non-completion	Significant in K: 1/5 Non-significant in K: 4/5 Direction: positive N: - Effect: -
	6	Traditional higher education	Ethnicity	Non-completion	Non-significant in K: 4/4 Direction: - N: - Effect: -
	6	Traditional higher education	Social class	Non-completion	Significant in K: 1/2 Non-significant in K: 1/2 Direction: students with a doctor-parent are less likely to drop out N: - Effect: -

6	Traditional higher education	Parents' education	Non-completion	Significant in interaction with subgroup in K: 1/3 Non-significant in K: 2/3 Direction: fathers' education negatively related to dropout N: - Effect: -
8	Traditional higher education	Socioeconomic status	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Age	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Marital status	Completion	Significant Direction: diverse K: - N: - Effect: -
8	Traditional higher education	Parents' education	Completion	Significant Direction: positive K: - N: - Effect: -
10	Traditional higher education	Socioeconomic status	Completion	Significant in K: 1/1 Direction: positive N: - Effect: -

10	Dispositional cognitive	Traditional higher education	Gender	Completion	Significant in K: 2/6 Non-significant in K: 4/6 Direction: females show more persistence N: - Effect: -
2		Traditional higher education	High school GPA	Completion	Significant Direction: positive K: 12 N: 551 Effect: estimated true correlation = .246
2		Traditional higher education	ACT/SAT scores	Completion	Significant Direction: positive K: 11 N: 3053 Effect: estimated true correlation = .124
2		Traditional higher education	Academic-related skills	Completion	Significant Direction: positive K: 8 N: 1627 Effect: estimated true correlation = .366
9		Traditional higher education	Self-regulation	Completion	Non-significant Direction: - K: - N: - Effect: -

6	Traditional higher education	Entry qualifications	Completion	Significant in K: 9/12 Non-significant in K: 3/12 Direction: positive N: - Effect: -
8	Traditional higher education	Study habits	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Time-management skills	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Preparedness for course material	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Academic success	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Language	Completion	Significant Direction: students with the language of the study program as first language show higher retention rates K: - N: - Effect: -

5	Higher online education	Academic background: GPA	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Academic background: previous academic performance	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Academic background: SAT math score	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant experiences: educational level	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant experiences: number of courses completed online previously	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant experiences: experience in relevant field	Non-completion	Significant Direction: negative K: - N: - Effect: -

5	Higher online education	Relevant experiences: involvement in professional activities	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant skills: management skills	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant skills: resilience	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Relevant skills: computer skills	Non-completion	Significant Direction: negative K: - N: - Effect: -
10	Traditional higher education	Secondary school GPA	Completion	Significant in K: 16/16 Direction: positive N: - Effect: -
10	Traditional higher education	Secondary school mathematics GPA	Completion	Significant in K: 2/2 Direction: positive N: - Effect: -
10	Traditional higher education	Intelligence	Completion	Non-significant in K: 1/1 Direction: - N: - Effect: -

10	Traditional higher education	Prior education levels	Completion	Significant Direction: positive for university context; inconsistent for professional higher educational context K: - N: - Effect: -
10	Traditional higher education	Learning strategies: external regulation	Completion	Significant in K: 2/2 Direction: - N: - Effect: -
10	Traditional higher education	Learning strategies: lack of regulation	Completion	Significant Direction: negative K: - N: - Effect: -
10	Traditional higher education	Learning strategies: deep learning	Completion	Non-significant Direction: - K: - N: - Effect: -
10	Traditional higher education	Learning strategies: critical processing	Completion	Non-significant Direction: - K: - N: - Effect: -
10	Traditional higher education	Learning strategies: surface learning	Completion	Non-significant Direction: - K: - N: - Effect: -

10	Traditional higher education	Learning strategies: concrete processing	Completion	Non-significant Direction: - K: - N: - Effect: -
10	Traditional higher education	Learning strategies: conceptions of knowledge	Completion	Non-significant in K: 2/2 Direction: - N: - Effect: -
10	Traditional higher education	Learning skills preparation	Completion	Non-significant Direction: - K: - N: - Effect: -
Dispositional non-cognitive	2	Traditional higher education	Completion	Significant Direction: positive K: 7 N: 3208 Effect: estimated true correlation = .066
2	Traditional higher education	Academic goals	Completion	Significant Direction: positive K: 33 N: 20010 Effect: estimated true correlation = .340
2	Traditional higher education	Institutional commitment	Completion	Significant Direction: positive K: 28 N: 20741 Effect: estimated true correlation = .206

2	Traditional higher education	Social involvement	Completion	Significant Direction: positive K: 36 N: 26263 Effect: estimated true correlation = .216
2	Traditional higher education	Academic self-efficacy	Completion	Significant Direction: positive K: 6 N: 6930 Effect: estimated true correlation = .359
2	Traditional higher education	General self-concept	Completion	Non-significant Direction: positive K: 6 N: 4240 Effect: estimated true correlation = .050
2	Traditional higher education	Perceived social support	Completion	Significant Direction: positive K: 26 N: 11624 Effect: estimated true correlation = .257
9	Traditional higher education	Attributions	Completion	Non-significant Direction: - K: - N: - Effect: -

9	Traditional higher education	Anxiety	Completion	Non-significant Direction: - K: - N: - Effect: -
9	Traditional higher education	Motivation	Completion	Significant Direction: positive K: - N: - Effect: correlation = .150
9	Traditional higher education	Self-perceptions	Completion	Significant Direction: positive K: - N: - Effect: correlation = .100
6	Traditional higher education	Psychological variables	Non-completion	Significant in K: 3/3 Direction: - N: - Effect: -
4	Traditional higher education	Personality: Neuroticism	Completion	Non-significant Direction: - K: - N: - Effect: -
4	Traditional higher education	Personality: Extraversion	Completion	Non-significant Direction: - K: - N: - Effect: -

4	Traditional higher education	Personality: Openness to Experience	Completion	Non-significant Direction: - K: - N: - Effect: -
7	Traditional higher education	Adjustment to college	Completion	Direction: positive K: 23 N: 26299 Effect: true score correlation = .230
7	Traditional higher education	Academic adjustment	Completion	Direction: positive K: 13 N: 4116 Effect: true score correlation = .190
7	Traditional higher education	Social adjustment	Completion	Direction: positive K: 11 N: 3672 Effect: true score correlation = .250
7	Traditional higher education	Personal-emotional adjustment	Completion	Direction: positive K: 13 N: 4060 Effect: true score correlation = .130
7	Traditional higher education	Institutional adjustment	Completion	Direction: positive K: 13 N: 4371 Effect: true score correlation = .290
8	Traditional higher education	Relevance of program to goals and interests	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Perceived value of educational program	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Difficulties juggling commitments	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Student receptivity to available services	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Self-motivation	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Self-determination	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Student self-efficacy	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Goals and intentions	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Intention to gain knowledge and skills rather than to obtain degree	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Psychological needs met	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Sense of belonging	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Peer self-esteem	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Personality	Completion	Significant Direction: high psychoticism and extraversion are negatively related to retention K: - N: - Effect: -

8	Traditional higher education	Campus involvement	Completion	Significant Direction: positive K: - N: - Effect: -
5	Higher online education	Computer confidence	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Psychological attributes: locus of control	Non-completion	Significant Direction: internal locus of control is negatively related to drop out K: - N: - Effect: -
5	Higher online education	Psychological attributes: motivation	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Psychological attributes: goal commitment	Non-completion	x
5	Higher online education	Psychological attributes: love of learning	Non-completion	x
5	Higher online education	Psychological attributes: self-efficacy	Non-completion	Significant Direction: negative K: - N: - Effect: -

5	Higher online education	Psychological attributes: satisfaction	Non-completion	Significant Direction: negative K: - N: - Effect: -
5	Higher online education	Interactions: student-content	Non-completion	Significant Direction: negative K: - N: - Effect: -
10	Traditional higher education	Personality	Completion	Significant for conscientiousness Non-significant for neuroticism, extraversion, openness to experience and agreeableness Direction: negative K: - N: - Effect: -
10	Traditional higher education	Motivational characteristics: self-efficacy	Completion	Significant in K: 3/3 Direction: positive N: - Effect: -
10	Traditional higher education	Motivational characteristics: reasons for engagement, intrinsic motivation	Completion	Significant in K: 2/6 Non-significant in K: 4/6 Direction: positive N: - Effect: -
10	Traditional higher education	Motivational characteristics: reasons for engagement, extrinsic motivation	Completion	Non-significant Direction: - K: - N: - Effect: -

10	Traditional higher education	Motivational characteristics: reasons for engagement, study motivation	Completion	Significant in K: 2/3 Non-significant in K: 1/3 Direction: positive N: - Effect: -
10	Traditional higher education	Motivational characteristics: reasons for engagement, motivation to be involved in extracurricular activities	Completion	Non-significant Direction: - K: - N: - Effect: -
10	Traditional higher education	Motivational characteristics: reasons for engagement, lack of motivation	Completion	Significant in K: 2/2 Direction: negative N: - Effect: -
10	Traditional higher education	Psychological factors: academic adjustment, institutional attachment	Completion	Significant in K: 2/2 Direction: positive N: - Effect: -
10	Traditional higher education	Psychological factors: academic integration	Completion	Significant in K: 1/1 Direction: positive N: - Effect: -
10	Traditional higher education	Psychological factors: degree programme satisfaction	Completion	Significant in K: 2 Direction: positive N: - Effect: -
10	Traditional higher education	Behavioural engagement: Attendance	Completion	Significant Direction: positive K: - N: - Effect: -

10	Traditional higher education	Behavioural engagement: Observed learning activities, regular study behaviour	Completion	Significant in K: 2/2 Direction: positive N: - Effect: -
10	Traditional higher education	Behavioural engagement: self-study time	Completion	Significant in K: 2/2 Direction: positive N: - Effect: -
Situational	3	Traditional higher education	Employment	Significant Direction: 1-15 employment hours lowest risk for enrolment interruption, followed by 16-34 employment hours and no employment, 35+ employment hours at highest risk. K: 1/1 N: - Effect: -
			Completion	Significant Direction: positive K: 1/1 N: - Effect: -
				Significant in interaction with accommodation of work Direction: working on campus more likely to persist than working off campus K: 1/1 N: - Effect: -

5	Higher online education	Work commitments	Non-completion	Significant in K: 5/5 Direction: positive N: - Effect: -
5	Higher online education	Study environments: emotional support	Non-completion	Significant in K: 2 Direction: negative N: - Effect: -
5	Higher online education	Study environments: financial aid	Non-completion	Significant in K: 2 Direction: negative N: - Effect: -
5	Higher online education	Study environments: comfortable circumstances in which to study provided by environment	Non-completion	Significant in K: 3 Direction: negative N: - Effect: -
5	Higher online education	Study environments: life challenges experienced without sufficient support	Non-completion	Significant in K: 3 Direction: positive N: - Effect: -
5	Higher online education	Interactions: student-student	Non-completion	Non-significant in K: 2/2 Direction: - N: - Effect: -
8	Traditional higher education	Scholarship	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Financial difficulties	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Employment pressures	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Support networks	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Controlling parents	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Family responsibilities	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Experiences of racism	Completion	Significant Direction: negative K: - N: - Effect: -

8	Traditional higher education	Physical health	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	External pressures	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Substance use	Completion	Significant Direction: negative K: - N: - Effect: -
Institutional	1	Traditional higher education	Student-faculty informal contact	Completion
				Significant in K: 6/7 Non-significant in K: 1/7 Direction: positive N: - Effect: -
2	Traditional higher education	Financial support	Completion	Significant Direction: positive K: 8 N: 7800 Effect: estimated true correlation = .188
2	Traditional higher education	Size of the institution	Completion	Non-significant Direction: negative K: 6 N: 11482 Effect: -.010

2	Traditional higher education	Institution selectivity	Completion	Significant Direction: positive K: 6 N: 11482 Effect: .238
6	Traditional higher education	Educational initiatives	Non-completion	Significant in K: 1/2 Non-significant in K: 1/2 Direction: problem-based learning curriculum design less student dropout as compared to traditional curriculum design N: - Effect: -
5	Higher online education	Course design: team-building	Non-completion	Significant in K: 1/1 Direction: negative N: - Effect: -
5	Higher online education	Course design: well-structured, relevant content	Non-completion	Significant in K: 1/1 Direction: negative N: - Effect: -
5	Higher online education	Institutional support	Non-completion	Significant Direction: positive K: - N: - Effect: -
5	Higher online education	Interactions: faculty-student interaction	Non-completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Institutional size	Completion	Significant Direction: negative K: - N: - Effect: -
8	Traditional higher education	Long distance education programme availability	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Selectivity of the institution	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Clear course policies, procedures and guidelines	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Finances expended on student related services	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Cultural sensitivity	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Student-teacher relationships	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Mentors with clinical experience	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Availability of staff	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Timely feedback on assignments	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Transition programme to tertiary education	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	On campus support	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Orientation programme for students	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Orientation programme for parents	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Counselling services	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Vocational mentoring	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Availability of support programmes	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Student use of library services	Completion	Significant Direction: positive K: - N: - Effect: -

8	Traditional higher education	Peer mentoring	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Peer study groups	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Financial assistance	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Financial aid information availability	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Faculty availability	Completion	Significant Direction: positive K: - N: - Effect: -
8	Traditional higher education	Student advisors availability	Completion	Significant Direction: positive K: - N: - Effect: -

Note. ¹ For the exact terminology used for the outcome measure in the individual review studies we refer to the corresponding articles. For the sake of correct interpretation of the results (e.g., direction of relationships) we decided to indicate whether the outcome measure was formulated as completion (e.g., persistence, retention) or non-completion (e.g., drop-out or attrition). ² K = the number of studies/samples.

Appendix D – Results on interventions to raise completion rates in higher education (Chapter 1)

Intervention category	Ref	Education context	Intervention	Duration	Factor manipulated by the intervention (category)	N	Outcome measure	Effect on completion rates
Coaching (motivational) or remedial teaching (specific skills)	1	Traditional higher education	Access Plus program: a Student Success/Freshman Year Experience Office, an intrusive freshman advising program, an expanded Freshman Seminar course, redesigned Introduction to Professional Studies courses, enhanced orientation programs, creation of Freshman Interest Groups, added faculty for Developmental English and Mathematics, expansion of the programs of Center for Academic Support, the creation of a Unity Services Office, and a Center for Excellence in Teaching. (face-to-face)	x	Cognitive and non-cognitive factors amongst which academic motivation, social motivation, general coping skills and receptivity to institutional support as measured in a baseline by the College Student Inventory, but not measured after the intervention and not further specified. (dispositional cognitive and dispositional non-cognitive)	x	x	Significant: - Effect: 10% increase for freshmen to sophomore and 21% increase for at-risk students
	4	Traditional higher education	Optimism and Attributional Retraining (online (a)synchronous communication and face-to-face components)	30 minutes excl. informal session	Dispositional optimism Attributions (both dispositional non-cognitive)	Total: N = 236 Exp. group: N = 184 Control group: N = 52	Voluntary course withdrawal (number of credit hours dropped)	Significant for high-optimism students Effect: high-optimistic students who received AR

9	Traditional higher education	Developmental education (Texas Academic Skills Program) on math, reading and writing (face-to-face intervention)	x	Basic competences in a number of courses, such as mathematics and language skills (including writing) College preparedness (dispositional cognitive)	Two year colleges: Total: 255878 Experimental group: N = 104405 Control group: N = 151473 Four year colleges: Total: N = 197502 Experimental group: N = 44617 Control group: N = 152885	Probability of completing at least one year in college	training dropped fewer course hours than did those who did not receive AR training.
							Significant only when controlling for baseline covariates and only in 2 year colleges Effect: the estimates indicate that remediation for those starting at two-year colleges lowers the probability of completing at least one year in college by 6 percentage points.

11	Traditional higher education	Coaching program by a student coaching service (InsideTrack) (online intervention synchronous and asynchronous communication)	One year	Lack of access to appropriate information Students' academic preparation and performance Lack of integration into the university community (dispositional cognitive, dispositional non-cognitive, situational)	Exp. group: N=8049 Control group: N=5506 Total: N=13555	Persistence	Significant at the 1% level Effect: the persistence rate of the coached students was 63% compared to 58% for uncoached students at 6 months. The treatment effect is constant at about 5 percentage points. At 12 months, the persistence rate for coached students is 48.8%, compared to 43.5% persistence rate for uncoached students. The estimated effect represents a 12% increase in college persistence. After 18 months, the treatment effect was 4.3 percentage points, representing a 15% increase in persistence in this
----	------------------------------	--	----------	---	---	-------------	--

sample, and after 24 months, the treatment effect was 3.4 percentage points, representing a 14% increase in persistence. All these differences are significant at the 1% level and do not change when controlling covariates (e.g., age, SAT score).					
12	Traditional higher education	Remedial courses, consisting of 160h of lectures, covering both mathematics and language skills (face-to-face intervention)	2 months, 160 hours	Basic competences in a number of courses, such as mathematics and language skills (dispositional cognitive)	Probability of dropping out of university (students who do not have passed any exam over the first 2 years of their degree program are considered dropped out or at strong risk of dropping out) Total: N = 4019 Significant for the first year, non-significant in the second year. Effect: the remedial program effect implies a reduction in the dropout probability of about 25-54%.

13	Traditional higher education	Self-regulated learning course (face-to-face intervention)	x	Self-regulated learning (dispositional cognitive)	Total: N = 29319 Completed SRL course: N = 4981	Retention (a retention outcome variable for each year of the study was created by coding anyone who was enrolled in the subsequent fall of who had graduated by that fall as a 1, and coding all the others 0)	Significant at the 1% level when controlled for demographic and academic background variables. Effect: students who complete SRL course during their freshmen year are nearly twice as likely to be retained to the second year, regardless of background characteristics. The odds that SRL completers will succeed increase each year through the fifth year.
----	------------------------------	---	---	--	--	---	--

Peer mentoring	3	Traditional higher education	Peer mentoring (a combination of online synchronous communication, online asynchronous communication, and face-to-face components)	One semester	Academic and interpersonal skills (dispositional cognitive and dispositional non-cognitive)	Experimental group: N=53 Total: N=53	Academic standing (coded as retained in good academic standing, retained on warning, retained on probation and academically dismissed, based on GPA)	Significant at the 5% level Effect: before the intervention, all mentees were on academic warning or academic probation. After the intervention 23/53 mentees retained in good academic standing, 3 retained on warning, 6 on probation and 21 students were academically dismissed.
	5	Traditional higher education	Peer mentoring (face-to-face intervention)	One year	Academic and social integration (dispositional non-cognitive)	Total: N = 128 Experimental group: N = 56 Control group: N = 72	Academic status (good standing, on probation or required to withdraw from the first to second year)	Significant at the 1% level Effect round 1: 88,5% retention in the mentored group, 57,1% in the control group. Round 2: 71,4% retention in the mentored group, 23,1% in the control group.

8	Traditional higher education	Peer mentoring (face-to-face intervention)	One year	College adjustment Motivation Career decision (all three dispositional non-cognitive)	Experimental group: N = 150 Control group: N = 157	Persistence (persistent = reregistering for a MST program after one year of college, non-persistent = reregistering for a non-MST program or dropping out of college altogether)	Significant at the 5% level Effect: mentees of the MIREs mentoring program were more likely to persist (86%) than did control group participants (76%).
Motivational contact	15	Higher online education	Approx. 6 months	Motivation (dispositional non-cognitive)	Experimental group: N = 1683 Control group: N = 1691	Retention (entered at least 1 exam, sat at least 1 exam, or passed at least 1 exam)	Significant at the 10% level Effect: 2.3 percentage points increase in retention.
	6	Higher online education and traditional higher education	One course or semester	ARCS factors (dispositional non-cognitive)	Total: N = 153 Exp. online group: N = 63 Exp. face-to-face group: N = 32 Control (online) group: N = 58	ARCS constructs (CIS survey) Withdrawal (failure at the course test and withdrawal together classified as non-completers)	Significant at the 5% level between the control (online) group and the experimental online group. No statistically significant difference between experimental online group and

experimental face-to-face group.

Effect: 34,5% non-completer rate in the control (online) group and 11,1% non-completer rate in the experimental online group.

Significant: -

Effect: the first intervention (by telephone only) increased retention by around 5%. The second intervention (by telephone, email and letters) increased retention by around 25 percentage points.

7	Higher online education	Proactive motivational contact by telephone, email and letters (online synchronous and online asynchronous communication intervention)	One course	Motivation (dispositional non-cognitive)	First intervention N = 987-10130 Second intervention Experiment al group: N = 13-15 Control group: N = 398-403	Completing the course	Significant: - Effect: the first intervention (by telephone only) increased retention by around 5%. The second intervention (by telephone, email and letters) increased retention by around 25 percentage points.	
10**	Academic dismissal policy	Traditional higher education	Academic dismissal policy (warning, providing extra support and binding study advice)	One year	1. Efficient use of study time, in this article referring to initial aptitude, academic procrastination, motivation and adaptation to the academic environment 2. Identification of students	Experiment al group: N=407 (2005) and N=402 (2006) Control	Dropout rate (the proportion of students from initial cohorts who left medical	Significant at the 1% level for dropout rate measured 2 years after enrolment and not significant for

	who are experiencing academic difficulties and providing support	group: N=406 (2003) and N=403 (2004) Total: N=1618	school voluntarily or were dismissed measured after 5 months and at 2 years after enrolment)	dropout rate at 5 months and for completion of year 1 curriculum at 2 years after enrolment.
(online asynchronous communication and face-to-face components intervention)	3. Curriculum-related or policy-related aspects			
	1. (dispositional cognitive and dispositional non-cognitive) 2. and 3. (institutional)			
			Year 1 curriculum completion rate (proportion of students from the initial cohorts who earned all 60 credits in the Year 1 curriculum and was measured at 2 years after enrolment)	Effect: at 2 years after enrolment the dropout rate was significantly higher in the academic dismissal condition, compared to the non-academic dismissal condition with an effect size of 0.07 calculated directly from chi-squared tests.
Traditional higher education	Curriculum-related or policy-related aspects (institutional)	Total: For 450 bachelor programs and 6 years the total N = 2700.	Reregistering (number of students reregistering after their first bachelor year)	Significant at the 1% level for all domains except for engineering (significant at the 5% level), medicine (not significant), liberal arts (not significant) and dropout rate
Academic dismissal policy (binding study advice)				
(online asynchronous communication intervention)				

16 **	Traditional higher education	Academic dismissal policy (online asynchronous communication intervention)	One year	Curriculum-related or policy-related aspects (institutional)	Exp.al group: N=1486 unique programs Control group: N = 266 unique programs Total: 6012 higher education degrees divided over 4 years.	Dropout (percentage of full-time bachelor students who leave the institution during the first year of enrolment)	((1 - proportion reregistering)/the number of enrolled students in the first year) Effect: overall the 1-year dropout rate is 0.358 for binding study advice students, compared to 0.279 for students in a program not providing binding study advice. Programs providing the binding study advice show on average 7.1% higher first-year dropout. Significant at the 0.01% level. Effect: more students drop out in the first year in the programs in which there is an academic dismissal policy. Dropout will increase by 7.5% by the introduction of an academic dismissal policy.
----------	------------------------------	---	----------	---	---	--	---

Intervention on instruction	2	Higher online education	A list of diverse systematic instructional methods in online courses for new online students (online synchronous communication and online asynchronous communication intervention)	One course or semester	ARCS factors (dispositional non-cognitive)	Filled in the measures: N = 12 - 20 Not clear on how many students the figures on retention are based	ARCS constructs post-measure-ment Retention	Significant: - Effect: between fall 1989 and 1996 fall 44% of the students dropped out of the program by their third course. At the end of 1997 (after 3 semesters of the intervention) the attrition rate decreased to 22%, by the end of 1999-2000 academic year, the attrition rate decreased to 15%.
-----------------------------	---	-------------------------	---	------------------------	---	---	---	---

Note. ** These articles concern the same intervention for overlapping data sets. Article 14 is about Dutch university samples from 2002-2007, article 16 is about Dutch higher education samples (including university samples) from 2003-2004 and 2008-2009 and article 10 is about a specific single Dutch university sample from 2003-2004 and 2005-2006.

Appendix E – Questions and questionnaires (Chapter 2)

Variable	Instruction	Factor(s)	Items	Answer options
Academic self-efficacy	The next set of questions is about academic self-efficacy. Several study-related activities and achievements will be presented. To what extent do you estimate yourself capable of the following tasks or achievements?	Confidence in basic study skills	1. Take multiple choice exams	F1 (completely unable) to 5 (completely able)
			2. Work in an electronic learning environment	
			3. Understand most ideas presented in a text	
			4. Solve simple mathematical calculations	
			5. Make connections between different courses	
			6. Apply studied material in an assignment	
Hours planned to study		Discipline confidence	1. Study regularly	1 (completely unable) to 5 (completely able)
			2. Study regularly, instead of last minute	
			How many hours do you expect to study on average per week?	Open ended
Employment¹			1. Are you employed?	a) Yes, full time b) Yes, part time c) No, I am not employed
			2. What kind of employment contract do you have? Multiple answers possible.	a) Regular (paid) contract b) Freelance c) Volunteer
Basic mathematical skills	This test includes several tasks. Try to solve these tasks as well as possible. Do not use a calculator, mobile phone, or any other aid. You are allowed to write down calculations on a sheet of paper. Simplify your answer where possible. You do not have to display units (such as kilometres, seconds ...). A number is sufficient. For decimal numbers, you can use a comma or a dot. However, do not put spaces.	Basic mathematical skills	1. $8 - 0.8 * 5 =$	Open ended
			2. $3/5 * (-2) + 1/5 =$	Open ended
			3. A group of 400 people consists of 270 men and 130 women. The proportion of women with a low educational level is 0.4. How many women have a low educational level?	Open ended
			4. $(-1) - (-6) =$	Open ended
			5. Which of these sums is smaller than 1?	a) $1/2 + 5/9$ b) $7/8 + 1/4$ c) $2/3 + 4/12$ d) $2/5 + 1/4$ e) None of the above

		6. Give 1 correct answer: the square root of 0.01 =	Open ended
		7. $F(x)$ is the amount of gasoline in my car as a function of the travelled distance x (in kilometres travelled after a full refuelling). $F(x) = 50 - 0.05x$. After how many kilometres will my tank be empty?	Open ended
		8. Give 1 correct answer: solve the equation to x . $2x^2 + 4 = 3x^2 - 5$, so $x =$	Open ended
		9. $24 = 75\%$ of ...	a) 0.32 b) 18 c) 32 d) 36 e) None of the above
Goal orientation	The following statements relate to your goal to study. For each statement, indicate to the extent to which you agree with it.	Work avoidance	1. I want to do as little as possible in a course. 2. I want to complete a course with as little effort as possible. 3. I do not want to work hard in a course.
		Performance approach	1. Doing better than other students in a course is important to me. 2. It is important to me to do well compared to other students in a course. 3. My goal in a course is to get a better grade than most of the other students.
		Contact with other students	1. I try to work with other students from this course to complete the course assignments. 2. When studying for this course, I often set aside time to discuss course material with a group of students from the course. 3. I try to identify students in this course whom I can ask for help if necessary.
Learning strategies	Everyone has a certain way of studying. You too will notice that you will adopt a certain way of studying for the course(s) for which you recently registered. The following statements relate to the way in which you intend to study. For each statement, indicate to what extent you agree with it.	Discipline ²	1. I often do not finish what I planned to, because I feel lazy or tired. 2. I find it hard to stick to a study scheme. 3. I often find that I do not put enough time in a course, because of other activities.
			1 (completely disagree) to 7 (completely agree)

	Elaboration	1. I treat the course material as a starting point and I try to develop my own ideas about it.	1 (completely disagree) to 7 (completely agree)
		2. I try to relate ideas about a certain subject in a course to other courses.	
	Organization	3. I try to relate my own ideas to what I learn in a course.	1 (completely disagree) to 7 (completely agree)
		1. While I study for a course, I make an overview to organize my thoughts.	
		2. When I study for a course, I go over my notes and make an outline of important concepts.	
		3. When I study for a course, I write brief summaries of the most important ideas from the course material and my notes.	
Social support³	Indicate the extent to which you receive different types of support from different people in your environment.	1. How much financial support do you receive in regard to your study career?	1 (none at all) to 5 (very much), not applicable to me
		2. How much mental support do you receive in regard to your study career?	
		3. How much practical support do you receive in regard to your study career? For example: when someone supports you in quizzing, looks after your children while you are studying or in case someone helps you doing your assignments.	
		What is your goal in studying at the OUNL?	
Study intention		a) I want to complete one or more course(s)	
		b) I want to complete a short study program	
		c) I want to attain a bachelor degree	
		d) I want to attain a master degree	
		e) I want to complete a specific study program to be able to enrol in a master program afterwards	
		f) Orientation	
		g) I do not have a specific goal	

Volition	The following questions are aimed at getting an idea of how you expect to cope with unanticipated situations during your studies. We ask you to take the following situation as a starting point: if at any point I notice that I am not working in a targeted way and with little concentration, then ...	Consequence control	1. I picture the consequences of not continuously devoting myself to my goals/assignments/projects. 2. I think about the consequences of procrastinating my current study tasks too far ahead. 3. I think about the negative effects of not doing or finishing my tasks/projects. 4. I think about the consequences of moving my current tasks too far ahead.	1 (not applicable to me at all) to 5 (completely applicable to me)
		Metacognition	1. I use learning and working strategies, such as time management, to be able to continue working productively. 2. I reflect on my planning and adjust the associated goals if necessary. 3. I check the priorities in my current tasks. 4. I evaluate the feasibility of my learning goals.	1 (not applicable to me at all) to 5 (completely applicable to me)

Note. ¹ For every answer option respondents chose, they were asked to indicate the amount of hours they spend on this kind of employment contract. All hours in a paid employment context (regular (paid) contract and freelance) were added to calculate a new variable, in which participants indicating not to be employed (or only as a volunteer) were given value 0. ² Items were reverse coded before calculating factor scores. ³ All these questions were asked for support from different sources: partner, family, employer, colleagues, and friends. A value 0 means no support was received (or 'not applicable') from either of these sources in a type of support. A value 1 means support was received from at least one of the sources in a type of support.

Appendix F – Factor analyses process and cut-off values (Chapter 2)

Analysis	Stage of analysis	Setting	Cut-off value	Reference(s)
EFA	Assumptions	Barlett's test of sphericity	Should be significant	Field (2009)
		KMO measure of sampling adequacy for separate items	≥ 0.500	Field (2009)
	Method of factor extraction	Principal axis factoring	-	Brown (2006)
	Number of factors based on	Parallel analysis	-	Costello & Osborne (2005), Field (2009), Hayton et al. (2004), Ledesma & Valero-Mora (2006)
		Eigenvalues	≥ 1	
		Scree plot	-	
	In-/excluding items based on	Factor loadings of individual items	≥ 0.400 and no meaningful loadings on more than one factor	Comrey & Lee (1992), Field (2009)
	Rotation of factors	Oblique	-	Theoretical considerations per scale
	Model fit	Chi-square statistic p-value	Desirably ≥ 0.05, but defensible in case < 0.05, as this is usually the case in larger sample sizes.	Chin et al. (2008)
		Root Mean Square Error of Approximation (RMSEA)	≤ 0.08	Byrne (2010)
E-CFA		Standardized Root Mean Square Residual (SRMR)	≤ 0.07	Hooper et al. (2008)
		Comparative Fit Index (CFI)	≥ 0.95	Hooper et al. (2008)
		Tucker-Lewis Index (TLI)	≥ 0.95	Hooper et al. (2008)
		Standardized residuals	≤ 2.58	Brown (2015)
	Modification indices ¹	Factor loadings modification indices	≤ 11	Brown (2015)
		Residual covariances modification indices ²	≤ 11	Brown (2015)

*Note.*¹ Modification indices were always taken into account in combination with theoretical and practical considerations, as suggested by Byrne (2010). We did not follow-up on modification indices; if it would not make sense theoretically or if it would violate our aim for the most simple and parsimonious model (also see Note 2). In case other modification indices (even though less extreme) would make more sense or the model fit was acceptable even though there were still some indications for modifications, we took those reasons into account first.² Residual covariances modification indices were taken into account first as we aimed for the most simple and parsimonious tests, and allowing items to load on more than one factor, or allowing a correlated error variance between items (as indicated by standardized residuals) would violate that aim more than removing items based on residual covariances modification indices. See Brown (2015) and Byrne (2010) for more information.



Summary



Summary

Over the past decade, online higher education has expanded (Allen & Seaman 2017; Seaman et al., 2018). Due to the Covid-19 pandemic, this trend further accelerated. Moreover, it is predicted that after the pandemic, forms of online education will continue to play a prominent and lasting role (Gomez Recio & Colella, 2020). The openness and flexibility of online delivery come at a price, as indicated by higher non-completion rates in *online* higher education as compared to traditional (face-to-face, full-time) education (Patterson & McFadden, 2009; Simpson, 2013; Vossensteyn et al., 2015).

To enhance retention in online higher education, it is advised to address student commitment in an early stage, even prior to enrolment (Muljana & Luo, 2019). To that end, pre-enrolment (online) self-assessments seem a promising approach to raise awareness and provide early remediation (Nolden et al., 2019). Such assessments are advisory and informative instruments, which are conducive to self-examination (Hornke et al., 2013). In these self-assessments, prospective students complete tests on knowledge, skills, and/or attitudes proven predictive of completion in (online) higher education (Nolden et al., 2019). They receive feedback and advice that enable a well-informed study decision. Next to that, this feedback and advice possibly lead to early remediation, all for a solid start and successful continuation in higher (online) education (Broos et al., 2018; 2019; Kubinger et al., 2012; O'Regan et al., 2016).

Given the possible impact for students (study decision) and institutions (e.g., enrolment rates), the development of such self-assessments requires thorough and careful validation. Not only of the instrument itself but also of the way it is used and whether it affects the decision-making and study preparation process as intended. After all, it potentially is an impactful intervention. The main objective of the research presented in this thesis is to evaluate five sources of validity evidence for the purpose of designing a self-assessment for informed study decisions in online higher education. In line with state-of-the-art validity theory (AERA et al., 2014; Beckman et al., 2005; Cook et al., 2014), the following five sources of validity evidence are evaluated: *content*, *predictive value*, *internal structure*, *response processes*, and *consequences*. With this objective, the lack of (explicit) empirical evaluation of such self-assessments is addressed (Niessen & Meijer, 2017). The sources of validity evidence are evaluated in three parts, according to design-based research stages of *Analysis*, *Design and Development*, and *Evaluation* (Van den Akker et al., 2013).

Analysis stage

Chapter 1 involves a literature review focusing on predictors of completion and interventions aimed at enhancing completion in higher (online education). We review 10 review articles on predictors of completion and 16 articles on interventions aimed at enhancing completion. Four categories of predictors of completion are established:

- Demographic or background factors (e.g., prior level of education)
- Dispositional cognitive factors (e.g., mathematical skills) and non-cognitive factors (e.g., goals and intentions)

- Situational factors (e.g., social support)
- Institutional factors (e.g., faculty-student interaction)

Predictors showed diversity with respect to their consistency in predicting completion across the included studies and their modifiability, which is the extent to which they can be influenced by interventions.

In regard to interventions, findings indicate that coaching, remedial teaching, and peer mentoring are promising interventions for enhancing completion rates in higher education. However, the extent to which interventions focus on relevant consistent predictors, as established in the first part of the review, is limited. Also, evaluations of interventions *prior* to student enrolment are underrepresented in scientific literature.

Though scientific attention for interventions *prior* to student enrolment is lacking, research emphasized a need for such early interventions (Muljana & Luo, 2019). In this regard, self-assessments prior to student enrolment seemed a promising approach to build resources for completion in an early stage (e.g., Nolden et al., 2019).

In **Chapter 2**, we take the first steps in designing such a self-assessment. The predictors as demonstrated in Chapter 1 are the starting point. Instruments to measure those predictors are evaluated on internal structure and predictive value for the specific context for which the self-assessment is developed, the Open University of the Netherlands. Factor analyses indicated that most internal structures differed from previous instrument validation, which mainly took place in the context of *traditional* higher education. The predictive value of factors was analysed via CIBER and classification analyses. Based on a sensitivity cut-off of 95% (to minimize false negatives) about 13% of the actual non-completers could be identified correctly by the significant predictors. Resulting from this study is the first step towards a prototypical self-assessment with subtests on the following predictors of completion: numerical skills, discipline, study goals (hours planned to study), and social support.

All in all, the first part of this thesis, resulted in an explanation of the non-completion problem in the sense of predictors of completion, their predictive consistency and modifiability. In addition, we provide insight into how to measure such predictors in a self-assessment *prior* to student enrolment. More specifically, we address the following validity aspects of self-assessment: content, internal structure, and predictive value.

Design and development stage

Chapter 3 focuses on the content aspect of validity from a user perspective. After all, if (potential) users do not perceive the self-assessment and its subtests to be relevant, the chance that they will deliberately engage with the self-assessment decreases. In the long run, this might have critical implications for the impact of the self-assessment in enhancing completion rates. Furthermore, so far for the content aspect of validity, we focused only on the content of the self-assessment and its subtests. The main purpose of the self-assessment is to provide food for thought and feedback for action. Hence, the

content of the feedback aligned to the subtests requires a systematic and well-considered design process, too. Also in this regard, it is important not to ignore the perspective of (potential) users.

This study involves a survey about expectations of the self-assessment and feedback. Results from 66 prospective students indicate that not all prototypical subtests (Chapter 2) were considered relevant by prospective students. The fact that a numerical skills test was not considered as relevant, seemed to be an artefact of the sample (prospective students of law or cultural sciences). Social support was also perceived as less relevant, even though literature suggests that this is a highly relevant factor, especially in distance education (Muljana & Luo, 2019; Asbee & Simpson, 1998). Moreover, students rated eleven additionally suggested tests as relevant in the study decision process, indicating that the self-assessment might benefit from an expansion of tests.

In regard to the feedback, prospective students' expect an explanation of the measurement, advice for further preparation and odds for completing a course or program. A comparison of their obtained score to the score of a reference group was not expected. The results on feedback expectations are in line with the literature on learning analytic dashboards, which are often implemented for purposes similar to the self-assessment. Jivet and colleagues (2020), for instance, have shown that transparency (i.e., explanations of the scales used) and support for action (i.e., recommendations on how to change study behaviour) are important for students to make sense of such learning analytic dashboards and for self-regulated learning.

Based on these findings, the prototypical self-assessment is further developed, with a focus on the feedback aligned to the subtests. Three categories of feedback information are established: *My Score* (including a visualization indicating 'high', 'medium' and 'low-risk' scores), *Measurement* (i.e., explanation of the subtest), and *Advice* (i.e., general tips and links to resources and support services). Study advisors were closely involved in this process, as prospective students might contact them for further clarification or advice in following up the feedback.

In **Chapter 4**, a qualitative in-depth analysis is conducted, as a final step before 'full release' of the self-assessment. Eight prospective students took the self-assessment in an observed think-aloud mode. This provides insight into how they proceed through the self-assessment and what thought processes are elicited in order to fill out the tests, interpret the obtained scores and, determine possible steps in following up the feedback. Thereby, this study sheds light on the process and consequential validity aspects of the self-assessment. Findings reveal different response processes, depending on the type of subtest. The numerical skills test (i.e., a cognitive test) appears to elicit more invalid test-taking strategies (e.g., rushing, guessing), as compared to non-cognitive tests (e.g., discipline). On the latter, prospective students tend to derive their answers by reflecting on previous study experiences. Results in regard to the consequential aspect of validity show that the self-assessment does elicit reflection in the sense that prospective students think about whether they could do more in order to prepare themselves. Also, they felt encouraged by the fact that the feedback taught them that they can control their level of preparedness. The results also indicate that the consequential aspect of validity must

be considered in the context of decision-making phases. Prospective students in this study were already quite certain about enrolling. They stated that this affected their intention to follow-up the feedback and the impact of the self-assessment on their study choice certainty. We argue that given participants' favourable scores on the self-assessment, confirming the enrolment choice is a valid consequence.

The design and development stage led to the inclusion of two additional tests in the self-assessment: study strategy use and expectations of studying in higher online education. So far, these tests are prototypical, meaning that internal structure and score cut-offs are still to be investigated - beyond the scope of this thesis. Hence, prospective students do not yet receive a personalized score on these subtests, but general feedback only (i.e., explaining the measurements and providing general advice for further orientation or preparation in relation to the measured features).

Evaluation stage

After the full release of the self-assessment, evaluation took place on a larger scale. In **Chapter 5** we describe the results of this evaluation study in which prospective students ($N=662$) took the self-assessment in an authentic situation of orienting towards studying at the OUNL. We examine the impact of self-assessment results on study choice certainty and enrolment decisions, thereby, providing further insight into the consequential validity aspect of the self-assessment. In this regard, we also look at the fairness of the self-assessment, i.e., the consistency of impact of the self-assessment across subgroups based on background variables (gender, age, prior level of education). Results showed that, in general, more favourable self-assessment results related to higher levels of study choice certainty, and (thereby) higher enrolment proportions. The self-assessment appears to have added value for a reasonably large group (68.9%). This was expressed, for example in the fact that prospective students tend to become more certain about enrolling if they do not obtain any 'high-risk' scores and – to a certain extent – seem to adequately downsize their certainty in case they do obtain such scores. These results demonstrate evidence for the consequential validity aspect, as they are in line with the purposes of the self-assessment. A threat to consequential validity appears to be the fact that women tend to be affected by obtaining 'high-risk' scores, whereas such a (valid) impact on men is lacking. Further research is needed to investigate whether and how the feedback provided to men should be adapted to ensure that they do not dismiss high-risk scores too easily. Finally, in contrast to what was found in Chapter 4 and in related research (Soppe et al., 2020), the impact of the self-assessment did not seem to depend on initial study choice certainty. The findings of Chapter 5 are in line with those by Van Klaveren et al. (2019), who found that the effect of risk (versus success) scores, presented in a study decision activity, was independent of being (overly) confident.

In the **General Discussion**, the results of the present thesis are discussed in light of the design-based research' aspired contributions to (potentially) various types of theories: *domain theories*, *design frameworks*, and *design methodologies* (cf. Edelson, 2002).

In the category of domain theories, the contributions to outcome theories on non-completion in higher (online) education and to context theories on study or career decision making are discussed. For outcome theories on completion in higher (online) education, the present thesis added to existing models by shedding light on the predictive consistency and modifiability of predictors of completion. Thereby, we highlighted relevant factors for future interventions to focus on, in tackling the non-completion problem in higher (online) education. In the present thesis, these predictors are translated into tests, that together, form a self-assessment for informed study decisions.

With regard to context theories, the self-assessment is viewed in the process of study or career decision making. We particularly focused on the role of choice certainty during this process. Based on previous studies, we expected that prospective students who are already quite certain about their decision would be mainly looking for confirmation of their decision. Though, in Chapter 5, we demonstrated that this is not necessarily the case. To some extent, prospective students adequately adjusted their certainty based on unfavourable self-assessment results.

In regard to design frameworks, guidelines for self-assessments for informed study decisions were derived from the results of the present thesis. The general discussion, for example, stresses how different contexts might ask for different design decisions for self-assessments. Where younger students in other educational contexts might appreciate their scores to be compared to successful students or other test-takers, the adult target group of the present self-assessment appears less interested in such comparisons.

Finally, for design methodology contributions, the integration of design-based research stages and state-of-the-art validity theories is discussed as a process model to develop self-assessments for informed study decisions. Thereby, the research in this thesis adds to the literature on (self-)assessments in the context of orientation and admission towards higher education, as such assessments are rarely thoroughly and explicitly validated. In addition, it adds to the validity literature, by providing a hands-on example of applied validation studies for *all* validity aspects.



Samenvatting



In het afgelopen decennium is de omvang van - en deelname aan online hoger onderwijs toegenomen (Allen & Seaman, 2017; Seaman et al., 2018). Als gevolg van de Covid-19 pandemie is deze trend verder versterkt. Er wordt voorspeld dat na de pandemie, online onderwijs een prominente en blijvende rol zal spelen (Gomez Recio & Colella, 2020). De toegankelijkheid en flexibiliteit van online onderwijs hebben echter een prijs, te zien aan de hogere mate van studie-uitval in vergelijking met traditioneel (fysiek, fulltime) onderwijs (Patterson & McFadden, 2009; Simpson, 2013; Vossensteyn et al., 2015).

Om studie-uitval in online hoger onderwijs te verminderen wordt geadviseerd om vroegtijdig te interveniëren, zelfs voor inschrijving (voor de poort) (Muljana & Luo, 2019). Daartoe lijken (online) zelftesten voor toekomstige studenten een veelbelovende aanpak, om bewustwording te creëren en tijdig te remediëren wanneer nodig (Nolden et al., 2019). In deze zelftesten vullen toekomstige studenten tests in over kennis, vaardigheden en/of attitudes die bewezen voorspellend zijn gebleken voor studie voortgang of uitval in het hoger (online) onderwijs (Nolden et al., 2019). Ze ontvangen daarop feedback en advies die een weloverwogen studiekeuze mogelijk maakt. Daarnaast leiden deze feedback en adviezen mogelijk tot vroegtijdige remediëring, hoe dan ook voor een gedegen start en succesvolle voortzetting in het hoger (online) onderwijs (Broos et al., 2018; 2019; Kubinger et al. 2012; O'Regan et al., 2016).

Gezien de mogelijke impact van zulke zelftesten voor studenten (de studiekeuze) en onderwijsinstututen (bijvoorbeeld de inschrijfpercentages), vereist de ontwikkeling van dergelijke testen een grondig en zorgvuldig validatieproces. Niet alleen van de test(en) zelf, maar ook van de manier waarop deze gebruikt wordt en invloed heeft op het beslisproces. Het is immers potentieel een belangrijke ingreep. Het hoofddoel van het onderzoek in dit proefschrift is om vijf bronnen van validiteitsbewijs te evalueren, ten behoeve van het ontwikkelen van een zelftest voor weloverwogen studiekeuzes in het online hoger onderwijs. In lijn met de state-of-the-art theorie over validiteit (AERA et al., 2014; Beckman et al., 2005; Cook et al., 2014) worden de volgende vijf bronnen van validiteitsbewijs geëvalueerd: inhoud, voorspellende waarde, interne structuur, responsprocessen, en consequenties. Met dit doel wordt het gebrek aan (expliciete) empirische evaluatie van dergelijke zelftesten aangepakt (Niessen & Meijer, 2017). De bronnen van validiteitsbewijs worden in drie fasen geëvalueerd, volgens de fasen van ontwerpgericht onderzoek (Van den Akker et al., 2013): *Analyse*, *Ontwerp/Ontwikkeling*, en *Evaluatie*.

Analyse fase

Hoofdstuk 1 betreft een literatuuronderzoek naar voorspellers van studie-uitval en interventies gericht op het verminderen van studie-uitval in het hoger (online) onderwijs. Er worden tien overzichtsartikelen over voorspellers van studie-uitval en zestien artikelen over interventies gesynthetiseerd. Vier categorieën van voorspellers van studie-uitval zijn vastgesteld:

- Demografische of achtergrondfactoren (bijvoorbeeld vooropleiding)
- Dispositionele (toe te kennen aan de student) cognitieve factoren (bijvoorbeeld numerieke vaardigheden) en niet-cognitieve factoren (bijvoorbeeld doelen en intenties)

- Situationele factoren (bijvoorbeeld sociale steun)
- Institutionele factoren (bijvoorbeeld de interactie tussen docenten en studenten)

De consistentie in voorspellende waarde van voorspellers bleek te verschillen. Evenals de modificeerbaarheid, zijnde de mate waarin ze kunnen worden beïnvloed door interventies.

Wat interventies betreft, wijzen de bevindingen uit dat coaching, remediërend onderwijs en peer mentoring veelbelovende interventies zijn voor het verminderen van studie-uitval in het hoger (online) onderwijs. De mate waarin de interventies zich richten op relevante consistente voorspellers, zoals vastgesteld in het eerste deel van dit literatuuronderzoek, is echter beperkt. Ook zijn evaluaties van interventies die zich afspelen voordat studenten zijn ingestroomd (voor de poort) ondervertegenwoordigd in de wetenschappelijke literatuur.

Hoewel wetenschappelijke aandacht voor interventies vóór inschrijving van studenten ontbreekt, benadrukt onderzoek wel een behoefte aan dergelijke vroegtijdige interventies (Muljana & Luo, 2019). In dit opzicht lijken zelftesten voor weloverwogen studiekeuzes een veelbelovende aanpak, om in een vroeg stadium te werken aan een succesvolle start en voortzetting in het hoger (online) onderwijs (bijv. Nolden et al., 2019).

In **Hoofdstuk 2** zetten we de eerste stappen in het ontwerpen van een dergelijke zelftest. De voorspellers, zoals aangetoond in hoofdstuk 1 vormen het uitgangspunt. Instrumenten om die voorspellers te meten worden geëvalueerd op interne structuur en voorspellende waarde voor de specifieke context waarvoor de zelftest ontwikkeld is, de Open Universiteit. Factoranalyses laten zien dat de interne structuur van de meeste instrumenten verschilt van eerdere validatie studies, die plaatsvonden in de context van het traditioneel hoger onderwijs. De voorspellende waarde blijkt uit de CIBER- en classificatie analyses. Op basis van een sensitiviteit van 95% (om vals-negatieve voorspellingen te minimaliseren) kan ongeveer 13% van de daadwerkelijke uitvallers correct worden geïdentificeerd door de significante voorspellers. Het resultaat van deze studie is een eerste stap in de richting van een prototypische zelftest met daarin de volgende deeltesten: numerieke vaardigheden, discipline, studiedoelen (uren van plan aan de studie te besteden) en sociale steun.

Al met al heeft het eerste deel van dit proefschrift geresulteerd in een verklaring van het studie-uitval probleem in de zin van voorspellers van studie-uitval, de consistentie van de voorspellende waardes en de modificeerbaarheid. Daarnaast geven we inzicht in hoe dergelijke voorspellers gemeten kunnen worden in een zelftest voorafgaand aan inschrijving van studenten. Daarmee geven we specifiek inzicht in de volgende validiteitsaspecten: inhoud, interne structuur en voorspellende waarde van de zelftest.

Ontwerp- en ontwikkelfase

Hoofdstuk 3 richt zich op het validiteitsaspect 'inhoud', vanuit een gebruikers-perspectief. Immers, als (potentiële) gebruikers de zelftest en deeltesten daarvan niet als

relevant zien, is de kans gering dat ze er gebruik van zouden maken. Op lange termijn kan dit kritieke gevolgen hebben voor de effectiviteit van de zelftest in het verminderen van studie-uitval. Verder hebben we ons voor het validiteitsaspect 'inhoud' tot nu toe enkel gericht op de inhoud van de zelftest en deeltesten. Echter, het belangrijkste doel van de zelftest is om aan te zetten tot reflectie op de studiekeuze en mate van voorbereiding. Het is daarom ook vereist dat de feedback volgend op de deeltesten gebaseerd is op een systematisch en weloverwogen ontwerpproces. Ook daarin is het belangrijk om het perspectief van de (potentiële) gebruikers niet te negeren.

Dit onderzoek betreft een enquête over verwachtingen van de zelftest en feedback. De resultaten van 66 toekomstige studenten van de Open Universiteit laten zien dat niet alle prototypische deeltesten (zie hoofdstuk 2) als relevant worden gezien. Het feit dat zij numerieke vaardigheden niet als relevant beschouwen lijkt toe te kennen te zijn aan de specifieke doelgroep in dit onderzoek (toekomstige studenten van rechtswetenschappen en cultuurwetenschappen). Ook de deeltest over sociale steun werd als minder relevant beschouwd, ondanks dat dit volgens de literatuur een relevante factor is, juist in het afstandsonderwijs (Muljana & Luo, 2019; Asbee & Simpson, 1998). Daarnaast beoordelen toekomstige studenten in dit onderzoek elf aanvullende testen als relevant voor hun studiekeuzeproces, wat erop wijst dat de zelftest baat zou hebben bij een uitbreiding van deeltests.

Betreffende de feedback verwachten toekomstige studenten een toelichting op wat er gemeten is, een advies voor verdere voorbereiding en een indicatie van hun slaagkans voor een cursus of opleiding. Een vergelijking van de door hun behaalde score met de score van een referentiegroep verwachten ze niet. Deze resultaten zijn in lijn met de literatuur over *learning analytic dashboards*, die vaak worden geïmplementeerd voor dezelfde doeleinden als de zelftest. Jivet en collega's (2020) toonden bijvoorbeeld aan dat transparantie (uitleg over gebruikte instrumenten) en aanbevelingen voor (vervolg) studiegedrag belangrijk zijn voor studenten om effectief met dit soort dashboards omweg te kunnen.

Op basis van de bevindingen in hoofdstuk 3 werd de prototypische zelftest verder ontwikkeld, met name de feedback gerelateerd aan de deeltesten. Drie categorieën van feedbackinformatie zijn ontworpen: '*Mijn Score*' (inclusief een visualisatie die aangeeft of de score een hoog, gemiddeld of laag risico op studie-uitval impliceert), '*Instrument*' (uitleg over wat er werd gemeten en waarom), en '*Advies*' (algemene tips voor verdere voorbereiding en verwijzingen naar meer ondersteuning). In dit proces werden studieadviseurs nauw betrokken, aangezien toekomstige studenten hen zullen contacteren voor verdere verduidelijking of advies bij het opvolgen van de feedback.

In **Hoofdstuk 4** wordt een diepgaand kwalitatief onderzoek beschreven, als laatste stap voordat de zelftest voor het brede publiek geïmplementeerd wordt. Acht toekomstige studenten gebruikten de zelftest in een hardop-denken modus. Dit geeft inzicht in hoe zij de zelftest doorlopen en welke denkprocessen worden uitgelokt bij het invullen van de tests, het interpreteren van de behaalde scores en het bepalen van

mogelijke vervolgstappen op basis van de feedback. Op die manier belicht dit onderzoek de volgende twee validiteitsaspecten: respons-processen en consequenties.

De bevindingen laten zien dat de respons-processen afhankelijk zijn van het type deelttest. De numerieke vaardigheden test (een cognitieve test) blijkt namelijk minder valide strategieën uit te lokken (bijvoorbeeld haasten of gokken) in vergelijking met niet-cognitieve testen (bijvoorbeeld de discipline deelttest). Bij dit laatste soort deelttesten zijn toekomstige studenten geneigd hun antwoorden af te leiden door te reflecteren op eerdere studie-ervaringen. Met betrekking tot het validiteitsaspect 'consequenties', blijkt dat de zelftest, zoals bedoeld, aanzet tot reflectie, in de zin dat toekomstige studenten nadenken over of zij zich nog beter zouden kunnen voorbereiden. Ook voelen ze zich gesterkt in hun zelfvertrouwen, omdat de feedback hen leert dat ze zelf controle hebben over de mate waarin ze voorbereid zijn. De resultaten van dit onderzoek geven ook aan dat de consequenties van de zelftest moeten worden bekeken in de context van verschillende fases in het studiekeuzeproces. Toekomstige studenten in dit onderzoek waren al vrij zeker van hun keuze om zich in te schrijven. Zij verklaarden dat dit van invloed was op hun intentie om zich nog verder voor te bereiden en op de impact van de zelftest op hun keuze. Wij beargumenteren dat dit, gezien de gunstige (lage risico) scores die zij haalden, een valide consequentie van de zelftest is.

De ontwerp- en ontwikkelfase heeft geleid tot een uitbreiding van deelttesten in de zelftest. Twee extra testen zijn toegevoegd: studie strategieën en verwachtingen van studeren in het online hoger onderwijs. Tot nu toe zijn deze testen prototypisch, wat betekent dat de interne structuur van deze instrumenten en de voorspellende waarde nog onderzocht moeten worden – buiten dit proefschrift. Toekomstige studenten krijgen op deze testen nog geen persoonlijke score, maar ontvangen algemene feedback. Dat wil zeggen, een uitleg van de instrumenten en een algemeen advies voor verdere voorbereiding met betrekking tot deze factoren.

Evaluatiefase

Nadat de zelftest volledig geïmplementeerd is, beschrijft **Hoofdstuk 5** een evaluatie op grotere schaal. In dit onderzoek gebruiken toekomstige studenten ($N = 662$) de zelftest in een authentieke situatie van oriëntatie op studeren aan de Open Universiteit. We beschrijven het effect van de zelftest op de keuzezekerheid van toekomstige studenten en hun beslissing om zich al dan niet in te schrijven. Daarmee geven we inzicht op het validiteitsaspect 'consequenties'. We kijken ook naar de *fairness* van de test, dat wil zeggen de consistentie van effecten voor verschillende subgroepen, gebaseerd op achtergrondkenmerken (geslacht, leeftijd, en vooropleiding). Uit de resultaten blijkt dat gunstigere scores op de zelftest samenhangen met een hogere mate van zekerheid om in te schrijven, en (daardoor) hogere kans op inschrijving. De zelftest blijkt voor een vrij grote groep toekomstige studenten (68.9%) van toegevoegde waarde te zijn. Dit drukt zich bijvoorbeeld uit in het feit dat toekomstige studenten zekerder worden als zij geen hoge risico scores behalen en – tot op zekere hoogte – hun zekerheid adequaat afzwakken als zij wel zulke score behalen. Deze resultaten zijn in lijn met het doel van de zelftest en leveren daarmee bewijs voor het consequentiële validiteitsaspect. Een bedreiging voor dit validiteitsaspect werd gevonden in het feit dat vrouwen geneigd zijn

hun zekerheid af te zwakken naar aanleiding van hoge risico scores, terwijl een dergelijk (valide) effect bij mannen ontbreekt. Er is verder onderzoek nodig om na te gaan of en hoe de feedback aan mannen moet worden aangepast om ervoor te zorgen dat zij zulke hoge risico scores niet zomaar verwerpen. Tenslotte blijkt dat, in tegenstelling tot de verwachtingen op basis van hoofdstuk 4 en eerder onderzoek (Soppe et al., 2020), het effect van de zelftest niet afhankelijk is van hoe zeker toekomstige studenten op voorhand al van hun keuze om in te schrijven zijn. De bevindingen in hoofdstuk 5 zijn in lijn met die van Van Klaveren et al. (2019). Zij vonden dat het effect van risicoscores (versus successcores), gepresenteerd in een gelijkaardige oriëntatieactiviteit, onafhankelijk was van zekerheid.

In de **Algemene Discussie** worden de resultaten van dit proefschrift besproken in bijdragen van ontwerpgericht onderzoek, te categoriseren in drie soorten theorieën: domeintheorieën, ontwerprichtlijnen en ontwerpmethodologieën (cf. Edelson, 2002).

In de categorie van domeintheorieën worden de bijdragen onderscheiden als bijdragen aan uitkomst-gerelateerde theorieën (studie-uitval) of context-gerelateerde theorieën (studie- en loopbaanbeslissingen). Voor uitkomst-gerelateerde theorieën draagt dit proefschrift bij aan bestaande modellen waarin studie-uitval wordt voorspeld door inzicht te leveren in voorspellende consistentie en modificeerbaarheid van voorspellers. Daarmee maken we inzichtelijk waar toekomstige interventies om studie-uitval aan te pakken zich op zouden kunnen richten. In dit proefschrift zijn voorspellers vertaald in deeltesten die samen een zelftest vormen voor weloverwogen studiekeuzes.

Met betrekking tot context-gerelateerde theorieën wordt de zelftest bekeken als onderdeel van het studie- of loopbaankeuzeproces. We richtten ons in het bijzonder op de rol van keuzezekerheid in dit proces. Op basis van eerder onderzoek verwachtten we dat de toekomstige studenten die al vrij zeker zijn van hun keuze vooral op zoek zouden gaan naar bevestiging daarvan. In hoofdstuk 5 toonden we aan dat dit echter niet noodzakelijk het geval is. Tot op zekere hoogte stellen toekomstige studenten hun zekerheid adequaat bij op basis van minder gunstige zelftest resultaten.

Met betrekking tot richtlijnen voor ontwerp zijn in dit proefschrift enkele richtlijnen naar voren gekomen voor zelftesten gericht op studiekeuzes in het hoger (online) onderwijs. We benadrukken bijvoorbeeld hoe verschillende contexten kunnen vragen om verschillende ontwerpbeslissingen. Waar jongere studenten (adolescenten) het wellicht op prijs stellen als hun zelftest scores worden vergeleken met de scores van succesvolle studenten of andere testgebruikers, lijkt de volwassen doelgroep van de zelftest in dit proefschrift daar minder in geïnteresseerd.

Tenslotte de bijdragen voor wat betreft de ontwerpmethodologie. We bespreken de integratie van state-of-the-art validiteitstheorie in de fasen van ontwerpgericht onderzoek als een procesmodel om zelftesten voor studiekeuzes te ontwikkelen. Dit proefschrift draagt daarmee bij aan wetenschappelijk inzicht over zelftesten in de context van oriëntatie en toegang tot het hoger onderwijs, aangezien dergelijke zelftesten zelden grondig en expliciet gevalideerd worden. Daarnaast draagt het bij aan de validiteitsliteratuur, door een praktijkvoorbeeld te geven van toegepast validiteitsonderzoek voor *alle* validiteitsaspecten.



References



References

References reviewed in Chapter 1 are marked with * for the part on predictors and ** for the part on interventions.

- Abbasi, N., & Ghosh, S. (2020). Construction and Standardization of Examination Anxiety Scale for Adolescent Students. *International Journal of Assessment Tools in Education*, 7(4), 522-534. <https://doi.org/10.21449/ijate.793084>
- Alexander, P. A., Graham, S., & Harris, K. R. (1998). A perspective on strategy research: Progress & prospects. *Educational Psychology Review*, 10(2), 129-154. <https://doi.org/10.1023/A:1022185502996>
- Allen, J., Robbins, S. B., & Sawyer, R. (2009). Can measuring psychosocial factors promote college success? *Applied Measurement in Education*, 23(1), 1-22. <https://doi.org/10.1080/08957340903423503>
- Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States*. Babson Survey Research Group and Quahog Research Group, LLC. <https://files.eric.ed.gov/fulltext/ED541571.pdf>
- Allen, I. E., & Seaman, J. (2017). *Digital learning compass: Distance education enrollment report 2017*. Babson Survey Research Group, e-Literate, and WCET. <https://www.onlinelearningsurvey.com/reports/digitallearningcompassenrollment2017>
- Allen, I. E., Seaman, J., Pouline, R., & Straut, T. T. (2016). *Online report card: Tracking online education in the United States*. Babson Park, MA. <https://files.eric.ed.gov/fulltext/ED572777.pdf>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Psychological Association.
- Aragon, S. R., & Johnson, E. S. (2008). Factors influencing completion and noncompletion of community college online courses. *The American Journal of Distance Education*, 22(3), 146-158. <https://doi.org/10.1080/08923640802239962>
- **Arnold, I. J. (2015). The effectiveness of academic dismissal policies in Dutch university education: An empirical investigation. *Studies in Higher Education*, 40(6), 1068-1084. <http://dx.doi.org/10.1080/03075079.2013.858684>
- Asbee, D. S., & Simpson, O. (1998). Partners, families and friends: Student support of the closest kind. *Open Learning: The Journal of Open, Distance and e-Learning*, 13(3), 56-59. <https://doi.org/10.1080/0268051980130309>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1-14. https://doi.org/10.1207/s15327809jls1301_1
- Beckman, T. J., Cook et al., D. A., & Mandrekar, J. N. (2005). What is the validity evidence for assessments of clinical teaching? *Journal of General Internal Medicine*, 20(12), 1159-1164. <https://doi.org/10.1111/j.1525-1497.2005.0258.x>
- Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist*, 39(4), 243-253. https://doi.org/10.1207/s15326985ep3904_6
- **Bettinger, E. P., & Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, 36(1), 3-19. <https://doi.org/10.3102/0162373713500523>
- Beyer, S. (1998). Gender differences in causal attributions by college students of performance on course examinations. *Current Psychology: Developmental, Learning, Personality, Social*, 17(4), 346-358. <http://dx.doi.org/10.1007/s12144-998-1016-5>
- Blackboard Inc. (2021). *Blackboard Collaborate* [Computer software]. <https://www.blackboard.com/teaching-learning/collaboration-web-conferencing/blackboard-collaborate>
- Borghans, L., Golsteyn, B., & Stenberg, A. (2015). Does expert advice improve educational choice? *PLoS One*, 10(12): e0145378. <https://doi.org/10.1371/journal.pone.0145378>
- Boscardin, C., Fergus, K. B., Hellevig, B., & Hauer, K. E. (2018). Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Medical Teacher*, 40(8), 855-861. <https://doi.org/10.1080/0142159X.2017.1396306>
- *Bowles, T. V., & Brindle, K. A. (2017). Identifying facilitating factors and barriers to improving student retention rates in tertiary teaching courses: A systematic review. *Higher Education Research & Development*, 36(5), 903-919. <https://doi.org/10.1080/07294360.2016.1264927>

References

- Broos, T., Verbert, K., Langie, G., Van Soom, C., & De Laet, T. (2018). Multi-institutional positioning test feedback dashboard for aspiring students: Lessons learnt from a case study in Flanders. *Proceedings of the 8th International Conference on learning Analytics and Knowledge, Sydney*, 51-55.
<https://doi.org/10.1145/3170358.3170419>
- Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2019). Learning dashboards at scale: Early warning and overall first year experience. *Assessment & Evaluation in Higher Education*, 45(6), 1–20.
<https://doi.org/10.1080/02602938.2019.1689546>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.
- Bryer, J., Andrade, H., & Cleary, T. (n.d.). *Diagnostic Assessment and Achievement of College Skills*.
<https://daacs.net/>
- Byrne, B. M. (2010). Multivariate application series. *Structural Equation Modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge/Taylor & Francis Group.
- Callan, M. J., Kim, H., & Matthews, W. J. (2015). Age differences in social comparison tendency and personal relative deprivation. *Personality and Individual Differences*, 87, 196-199.
<https://doi.org/10.1016/j.paid.2015.08.003>
- Carroll, D., Ng, E., & Birch, D. (2009). Retention and progression of postgraduate business students: An Australian perspective. *Open Learning: The Journal of Open, Distance and e-Learning*, 24(3), 197-209.
<https://doi.org/10.1080/02680510903201599>
- Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education: A Journal of Educational Research and Practice*, 12(2).
<https://doi.org/10.26522/brocked.v12i2.38>
- Chin, W. W., Peterson, R. A., & Brown, S. P. (2008). Structural equation modeling in marketing: Some practical reminders. *Journal of Marketing Theory and Practice*, 16(4), 287-298.
<https://doi.org/10.2753/MTP1069-6679160402>
- **Chyung, S. Y. (2001). Systematic and systemic approaches to reducing attrition rates in online higher education. *American Journal of Distance Education*, 15(3), 36-49.
<https://doi.org/10.1080/08923640109527092>
- Cizek, G. J., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement*, 70(5), 732-743.
<https://doi.org/10.1177/0013164410379323>
- Cobern, W. W., & Adams, B. A. (2020). When interviewing: How many is enough? *International Journal of Assessment Tools in Education*, 7(1), 73-79. <https://dx.doi.org/10.21449/ijate.693217>
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, 13(1), 15-42. https://doi.org/10.1207/s15327809jls1301_2
- Comrey, L. A., & Lee, H. B. (1992). *A first course in factor analyses* (2nd ed.). Lawrence Erlbaum Associates.
- Cook et al., D. A., Zendejas, B., Hamstra, S. J., Hatala, R., & Brydges, R. (2014). What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Advances in Health Sciences Education*, 19(2), 233-250. <https://doi.org/10.1007/s10459-013-9458-4>
- Corno, L., & Kanfer, R. (1993). The role of volition in learning and performance. In L. Darling-Hammond (Ed.), *Review of Research in Education* (Vol. 21, pp. 301-341). F.E. Peacock Publishes.
- Corno, L. (1994). Student volition and education: Outcomes, influences and practices. In D. H. Schunk, & B. J. Zimmermann (Eds.), *Self-regulation of Learning and Performance: Issues and Educational Applications* (pp. 229-254). Erlbaum.
- Costello, A. B. & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation*, 10(7).
<https://pareonline.net/getvn.asp?v=10&n=7>
- Creswell, J. W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education Limited.
- *Credé, M., & Niehorster, S. (2012). Adjustment to college as measured by the student adaptation to college questionnaire: A quantitative review of its structure and relationships with correlates and consequences. *Educational Psychology Review*, 24(1), 133-165.
<https://doi.org/10.1007/s10648-011-9184-5>
- Cross, K.P. (1981). *Adults as learners: Increasing participation and facilitating learning*. Jossey-Bass.
- Crutzen, R. & Peters G.-J. Y. (2019). *The book of behavior change*. <https://a-bc.gitlab.io/bbc/>
- Deimann, M., & Bastiaens, T. (2010). The role of volition in distance education: An exploration of its capacities. *The International Review of Research in Open and Distributed Learning*, 11(1), 1-16.
<https://doi.org/10.19173/IRRODL.V11I1.778>

- Delnoij, L. E. C., Dirkx, K. J. H., Janssen, J. P. W., & Martens, R. L. (2020). Predicting and resolving non-completion in higher (online) education – A literature review. *Educational Research Review*, 29, 100313. <https://doi.org/10.1016/j.edurev.2020.100313>
- Delnoij, L. E. C., Janssen, J. P. W., Dirkx, K. J. H., Gijssels, H. J. M., De Groot, R. H. M., Neroni, J., De Bie, M., & Martens, R. L. (2021). Predicting completion: The road to informed study decisions in higher online education. *Frontiers in Education*, 6. <https://doi.org/10.3389/educ.2021.668922>
- Delnoij, L. E. C., Janssen, J. P. W., Dirkx, K. J. H., & Martens, R. L. (2020). Designing an online self-assessment for informed study decisions: The user perspective. In C. Alario-Hoyos, M. J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, & S. M. Dennerlein (Eds), *Lecture Notes in Computer Science: Vol. 12315. Addressing Global Challenges and Quality Education*. Springer. https://doi.org/10.1007/978-3-030-57717-9_6
- Delnoij, L. E. C., Janssen, J. P. W., Dirkx, K. J. H., & Martens, R. L. (2021). *Towards an online self-assessment for informed study decisions: A mixed-methods validation study* [Manuscript in review]. Faculty of Educational Sciences, Open Universiteit.
- Demulder, L., Donche, V., & Lacante, M. (2019). Keuze-instrument Columbus geeft leerlingen zicht op hun zelfbeeld, motivatie en studiestrategieën: Waarom zijn deze niet-cognitieve vaardigheden belangrijk? [Decision-instrument Columbus provides students insight into their self-image, motivation and study strategies: Why are these non-cognitive skills important?]. *Caleidoscoop* 3, 28- 37. [https://www.mijnvclb.be/downloads/caleidoscoop/jg31%203%2028\(keuzeinstrument\).pdf](https://www.mijnvclb.be/downloads/caleidoscoop/jg31%203%2028(keuzeinstrument).pdf)
- **De Paola, M., & Scoppa, V. (2014). The effectiveness of remedial courses in Italy: a fuzzy regression discontinuity design. *Journal of Population Economics*, 27(2), 365-386. <https://doi.org/10.1007/s00148-013-0466-8>
- Dijkstra, P., Kuyper, H., Van der Werf, G., Buunk, A. P., & Van der Zee, Y. G. (2008). Social comparison in the classroom: A review. *Review of Educational Research*, 78(4), 828-879. <https://doi.org/10.3102%2F0034654308321210>
- Di Stasio, V., & Solga, H. (2017). Education as social policy: An introduction. *Journal of European Social Policy*, 27(4), 313–319. <https://doi.org/10.1177/0958928717728712>
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>
- Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*, 37(9), 830-837. <https://doi.org/j.1365-2923.2003.01594.x>
- Downing, S. M., & Haladyna, T. M. (2004). Validity threats: Overcoming interference with proposed interpretations of assessment data. *Medical Education*, 38(3), 327-333. <https://doi.org/10.1046/j.1365-2923.2004.01777.x>
- Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years?. *Frontiers in Psychology*, 7. <https://doi.org/10.3389/fpsyg.2016.00508>
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69–106. <https://doi.org/10.1111/j.1529-1006.2004.00018.x>
- Edelson, D. C. (2002). Design research: What we learn when we engage in design. *The Journal of the Learning Sciences*, 11(1), 105–121. <http://www.jstor.org/stable/1466722>
- Eil, D., & Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114-38. <https://doi.org/10.1257/mic.3.2.114>
- Elffers, L. (2018). Studiesucces in het hbo: Het belang van een ketenbenadering [Study success in higher education: The importance of a chain approach] In F. Glastra, & D. Van Middelkoop (Eds.), *Studiesucces in het hoger onderwijs* [Study success in higher education, from efficiency to societal relevance] (pp. 141-164). Eburon.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, 70, 461– 475. <https://doi.org/10.1037/0022-3514.70.3.461>
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501–519. <https://doi.org/10.1037/0022-3514.80.3.501>
- Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, 7(3), 93-99. <https://doi.org/10.1016/j.afjem.2017.08.001>
- Essig, G. N., & Kelly, K. R. (2013). Comparison of the effectiveness of two assessment feedback models in reducing career indecision. *Journal of Career Assessment*, 21(4), 519-536. <https://doi.org/10.1177%2F1069072112475283>
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole Publishing.

References

- Expertise Centrum Onderwijs (ECO). (2021). *Target Group Identification*. Internal report Open Universiteit: unpublished.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117-140. <https://doi.org/10.1177/001872675400700202>
- Field, A. P. (2009). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). Sage.
- *Fong, C. J., Davis, C. W., Kim, Y. W., Marriott, L., & Kim, S. (2017). Psychosocial factors and community college student success: A meta-analytic investigation. *Review of Educational Research*, 87(2), 388-424. <https://doi.org/10.3102/0034654316653479>
- Fonteyne, L., Duyck, W. (2015). Vraag het aan SIMON! [Ask SIMON!]. *Thema Hoger Onderwijs*, 2, 56-60. <https://www.ethicalforum.be/sites/default/files/Vraag%20het%20SIMON.pdf>
- Fonteyne, L., De Fruyt, F., Dewulf, N., Duyck, W., Erauw, K., Goeminne, K., Lammertyn, J., Marchant, T., Moerkerke, B., Oosterlinck, T., & Rosseel, Y. (2015). Basic mathematics test predicts statistics achievement and overall first year academic success. *European Journal of Psychology of Education*, 30(1), 95-118. <https://doi.org/10.1007/s10212-014-0230-9>
- Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences*, 56, 34-48. <https://doi.org/10.1016/j.lindif.2017.05.003>
- Gaytan, J. (2013). Factors affecting student retention in online courses: Overcoming this critical problem. *Career and Technical Education Research*, 38(2), 145-155. <https://doi.org/10.5328/cter38.2.147>
- Germeijs, V., & Verschueren, K. (2006). High school students' career decision-making process: Development and validation of the study choice task inventory. *Journal of Career Assessment*, 14(4), 449-471. <https://doi.org/10.1177/1069072706286510>
- Germeijs, V., & Verschueren, K. (2007). High school students' career decision-making process: Consequences for choice implementation in higher education. *Journal of Vocational Behavior*, 70(2), 223-241. <https://doi.org/10.1016/j.jvb.2006.10.004>
- Gomez Recio, S., & Colella, C. (2020). *The world of higher education after covid-19*. <https://www.yerun.eu/wp-content/uploads/2020/07/YERUN-Covid-VFinal-OnlineSpread.pdf>
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement & Evaluation in Counseling & Development*, 36(3), 181-191. <https://doi.org/10.1080/07481756.2003.11909741>
- Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement* 66(6), 930-944. <https://doi.org/10.1177/0013164406288165>
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29(2), 75-91. <https://www.jstor.org/stable/30219811>
- Hachey, A. C., Wladis, C. W., & Conway, K. M. (2013). Balancing retention and access in online courses: Restricting enrollment. Is it worth the cost? *Journal of College Student Retention: Research, Theory & Practice*, 15(1), 9-36. <https://doi.org/10.2190/CS.15.1.b>
- Harackiewicz, J. M., Durik, A. M., Barron, K. E., Linnenbrink-Garcia, L., & Tauer, J. M. (2008). The role of achievement goals in the development of interest: Reciprocal relations between achievement goals, interest and performance. *Journal of Educational Psychology*, 100(1), 105-122. <https://doi.org/10.1037/0022-0663.100.1.105>
- Hartnett, M. (2016). The importance of motivation in online learning. In M. Harnett, (Ed.) *Motivation in online education* (pp. 5-32). Springer. https://doi.org/10.1007/978-981-10-0700-2_2
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, 7(2), 191-205. <http://dx.doi.org/10.1177/1094428104263675>
- Henderikx, M. A., Kreijns, K., & Kalz, M. (2017). Refining success and dropout in massive open online courses based on the intention-behavior gap. *Distance Education*, 38(3), 353-368. <https://doi.org/10.1080/01587919.2017.1369006>
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-80. <https://doi.org/10.21427/D7CF7R>
- Hornke, L., Wosnitza, M., & Bürger, K. (2013). Self-Assessment: Ideen, Hintergründe, Praxis und Evaluation [Self-assessment: Ideas, background, practice and evaluation]. *Wirtschaftspsychologie*, 1(15), 5-16.

- Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, 104(1), 48–73. <https://doi.org/10.1037/a0026223>
- **Huett, J. B., Kalinowski, K. E., Moller, L., & Huett, K. C. (2008). Improving the motivation and retention of online students through the use of ARCS-based e-mails. *The American Journal of Distance Education*, 22(3), 159–176. <https://doi.org/10.1080/08923640802224451>
- IBM Corporation. (2016). *IBM SPSS Statistics for Windows* (Version 24.0) [Computer software]. <https://www.ibm.com/products/spss-statistics>
- **Inkelaar, T., & Simpson, O. (2015). Challenging the 'distance education deficit' through 'motivational emails'. *Open Learning: The Journal of Open, Distance and e-Learning*, 30(2), 152–163. <http://dx.doi.org/10.1080/02680513.2015.1055718>
- Jivet, I. (2021). *The Dashboard That Loved Me: Designing adaptive learning analytics for self-regulated learning*. [PhD dissertation, Open Universiteit].
- Jivet I., Scheffel M., Drachsler H., Specht M. (2017) Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In É. Lavoué, H. Drachsler, K. Verbert, J. Broisin, M. Pérez-Sanagustín (Eds.), *Lecture Notes in Computer Science*, Vol. 10474. Data Driven Approaches in Digital Education. Springer. https://doi.org/10.1007/978-3-319-66610-5_7
- Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., & Drachsler, H. (2020). From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. *The Internet and Higher Education*, 47, 100758. <https://doi.org/10.1016/j.iheduc.2020.100758>
- Jug, R., Jiang, X. S., & Bean, S. M. (2019). Giving and receiving effective feedback: A review article and how-to guide. *Archives of Pathology & Laboratory Medicine*, 143(2), 244–250. <https://doi.org/10.5858/arpa.2018-0058-RA>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527. <https://doi.org/10.1177/1356389011410522>
- Kane, M. (2010). Validity and fairness. *Language Testing*, 27(2), 177–182. <https://doi.org/10.1177/0265532209349467>
- Kappe, F. R., & Knuiman, C. (2019). *Studiekeuzecheck* [Study choice check] [Infographic]. Inholland university of applied sciences. <https://www.inholland.nl/onderzoek/publicaties/studiekeuzecheck>
- Keller, J. M. (2008). An integrative theory of motivation, volition, and performance. *Technology, Instruction, Cognition, and Learning*, 6(2), 79–104.
- King, W. R., & He, J. (2006). A meta-analysis of the technology acceptance model. *Information & Management*, 43(6), 740–755. <https://doi.org/10.1016/j.im.2006.05.003>
- Knuiman, C., & Kappe, F. R. (2017). *Studiekeuzechecks: een casusbundel. Beschrijvingen van de vorm en invoering van de SKC bij tien opleidingen van vijf hogescholen*. [Study choice check: case study collection. Descriptions of the form and implementation of the study choice check at ten study programmes of five universities of applied sciences]. Lectoraat Studiesucces, Hogeschool Inholland, in opdracht van de Vereniging Hogescholen. <https://www.inholland.nl/onderzoek/publicaties/studiekeuzechecks-een-casusbundel>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20985448>
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, 45(3), 214–222. <https://doi.org/10.5014/ajot.45.3.214>
- Kreiter, C. D. (2016). A research agenda for establishing the validity of non-academic assessments of medical school applicants. *Advances in Health Sciences Education*, 21(5), 1081–1085. <https://doi.org/10.1007/s10459-016-9672-y>
- Kubinger, K. D., Frebort, M., & Müller, C. (2012). Self-assessment im rahmen der studienberatung: Möglichkeiten und Grenzen. [Self-assessment in the context of student counselling: possibilities and limits] In K. D. Kubinger, M. Frebort, L. Khorramdel, & L. Weitensfelder, (Eds.), *Self-Assessment: Theorie und Konzepte*, [Self-assessment: Theory and Concepts] (pp. 9–24). Pabst Science Publishers.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162. <https://doi.org/10.1037/0033-2909.127.1.162>
- Kutlu, O., & Yavuz, H. C. (2019). An effective way to provide item validity: Examining student response processes. *International Journal of Assessment Tools in Education*, 6(1), 9–24. <https://doi.org/10.21449/ijate.447780>

References

- **Larose, S., Cyrenne, D., Garceau, O., Harvey, M., Guay, F., Godin, F., Tarabulsy, G. M., & Deschênes, C. (2011). Academic mentoring and dropout prevention for students in math, science and technology. *Mentoring & Tutoring: Partnership in Learning*, 19(4), 419-439. <https://doi.org/10.1080/13611267.2011.622078>
- *Lee, Y., & Choi, J. (2011). A review of online course dropout research: Implications for practice and future research. *Educational Technology Research and Development*, 59(5), 593-618. <https://doi.org/10.1007/s11423-010-9177-y>
- Lee, Y., Choi, J., & Kim, T. (2013). Discriminating factors between completers of and dropouts from online learning courses. *British Journal of Educational Technology*, 44(2), 328-337. <https://doi.org/10.1111/j.1467-8535.2012.01306.x>
- Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical Assessment, Research & Evaluation*, 12(2), 1-11. <https://doi.org/10.7275/wjnc-nm63>
- Lent, R. W., & Brown, S. D. (2013). Social cognitive model of career self-management: Toward a unifying view of adaptive career behavior across the life span. *Journal of Counseling Psychology*, 60(4), 557. <https://doi.org/10.1037/a0033446>
- Levy, Y. (2007). Comparing dropouts and persistence in e-learning courses. *Computers & Education*, 48(2), 185-204. <https://doi.org/10.1016/j.compedu.2004.12.004>
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20(3), 975-978. <https://doi.org/10.2466/pr0.1967.20.3.975>
- Lui, A. M., Franklin Jr., D., Akhmedjanova, D., Gorgun, G., Bryer, J., Andrade, H. L., Cleary, T. (2018). Validity evidence of the internal structure of the DAACS self-regulated learning survey. *Future Review: International Journal of Transition, College, and Career Success*, 1(1), 1-18. <http://www.futureinstitute.us/wp-content/uploads/2019/03/Future-Review-Online-Article-1.1.pdf>
- Martens R. L. (2018). Onderzoek en onderwijs verbinden [aligning research and education] In F. Glastra, & D. Van Middelkoop (Ed.), *Studiesucces in het hoger onderwijs* [Study Success in higher education, from efficiency to societal relevance] (pp. 231-258). Eburon.
- **Martorell, P., & McFarlin Jr, I. (2011). Help or hindrance? The effects of college remediation on academic and labor market outcomes. *The Review of Economics and Statistics*, 93(2), 436-454. https://doi.org/10.1162/REST_a_00098
- Mayers, A. (2013). *Introduction to statistics and SPSS in psychology*. Pearson Higher Ed.
- McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, 67(3), 267-277. <https://doi.org/10.1016/j.jclinepi.2013.08.015>
- McCann, E., & Turner, J. E. (2004). Increasing student learning through volitional control. *Teachers College Record*, 106(9), 1695-1714. <https://psycnet.apa.org/doi/10.1111/j.1467-9620.2004.00401.x>
- McGhie, V. (2017). Entering university studies: Identifying enabling factors for a successful transition from school to university. *Higher Education*, 73(3), 407-422. <https://doi.org/10.1007/s10734-016-0100-2>
- McGrath, C., Henham, M., Corbett, A., Durazzi, N., Frearson, M., Janta, B., Kamphuis, B. W., Katashiro, E., Brankovic, N., Guerin, B., Manville, C., Schwartz, I., & Schweppenstedde, D. (2014). *Higher education entrance qualifications and exams in Europe: A comparison*. http://www.rand.org/pubs/research_reports/RR574
- McKeachie, W. J., Pintrich, P. R., Lin, Y. G., Smith, D. A. F., & Sharma, R. (1990). *Teaching and learning in the college classroom. A review of the research literature*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge.
- Menon, M. E. (2004). Information search as an indication of rationality in student choice of higher education. *Education Economics*, 12(3), 267-283. <https://doi.org/10.1080/0964529042000258617>
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of Assessment. *Educational Researcher*, 18(2), 5-11. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/j.2330-8516.1988.tb00303.x>
- Messick, S. (1990). Validity of test interpretation and use. *ETS Research Report Series*, 1990(1), 1487-1495.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Microsoft Corporation. (2018). *Microsoft Visio* <https://products.office.com/en/visio/flowchart-software>.
- Mittendorff, K. (2015). *Kwalitatief evaluatieonderzoek naar de tevredenheid over, werking en resultaten van de intakeprocedure van Saxion* [Qualitative evaluation research into the satisfaction about, functioning and results of Saxion's intake procedure]. Internal Saxion report: unpublished.

- Morris, L. V., Wu, S. S., & Finnegan, C. L. (2005). Predicting retention in online general education courses. *The American Journal of Distance Education*, 19(1), 23-36. https://doi.org/10.1207/s15389286ajde1901_3
- Muljana, P. S., & Luo, T. (2019). Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review. *Journal of Information Technology Education: Research*, 18, 19-57. <https://doi.org/10.28945/4182>
- Neuville, S., Frenay, M., Schmitz, J., Boudrenghien, G., Noël, B., & Wertz, V. (2007). Tinto's theoretical perspective and expectancy-value paradigm: A confrontation to explain freshmen's academic achievement. *Psychologica Belgica*, 47(1). <http://dx.doi.org/10.5334/pb-47-1-31>
- Neroni, J., Gijsselaers, H. J. M., Kirschner, P. A., & De Groot, R. H. M. (2015). The Adult Learning Open University Determinants (ALoud) study: Biological and psychological factors associated with learning performance in adult distance education. *British Journal of Educational Technology*, 46(5), 953-960. <https://doi.org/10.1111/bjet.12288>
- Neroni, J., Meijs, C., Leontjevas, R., Kirschner, P. A., & De Groot, R. H. (2018). Goal orientation and academic performance in adult distance education. *International Review of Research in Open and Distributed Learning*, 19(2). <https://doi.org/10.19173/irrodl.v19i2.3440>
- Nicol, D. (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. *Assessment & Evaluation in Higher Education*, 34(3), 335-352. <https://doi.org/10.1080/02602930802255139>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218. <https://doi.org/10.1080/03075070600572090>
- Niessen, A. S. M. (2018). *New rules, new tools: Predicting academic achievement in college admissions*. [Doctoral dissertation, Rijksuniversiteit Groningen].
- Niessen, A. S. M., & Meijer, R. R. (2017). Voorspellen in het kader van de studiekeuzecheck: Tijd voor verbetering. [Predicting the study choice check: Time for improvement][Dutch] *Onderzoek van Onderwijs*, 46, 5-7.
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Predicting performance in higher education using proximal predictors. *PloS One*, 11(4). <https://doi.org/10.1371/journal.pone.0153663>
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Measuring non-cognitive predictors in high-stakes contexts: The effect of self-presentation on self-report instruments used in admission to higher education. *Personality & Individual Differences* 106, 183-189. <http://dx.doi.org/10.1016/j.paid.2016.11.014>
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2018). Admission testing for higher education: A multi-cohort study on the validity of high-fidelity curriculum-sampling tests. *PloS One*, 13(6). <https://doi.org/10.1371/journal.pone.0198746>
- Nisbet, I. (2019). Fairness takes centre stage. *Assessment in Education: Principles, Policy & Practice*, 26(1), 111-117. <https://doi-org.ezproxy.elib11.ub.unimaas.nl/10.1080/0969594X.2017.1358151>
- Nolden, P., & Wosnitza, M. (2016). Webgestützte selbstreflexion von abbruchrisiken bei Studierenden. [Web-based self-reflection of drop-out risk among students]. *Empirische Pädagogik*, 30(3/4), 576-603.
- Nolden, P., Wosnitza, M., Karabenick, S. A., Peixoto, F., Gonida, E., Stepanovic Ilic, I., Almeida, L. S., Stamovlasis, D., Taveira, M. D. C., Toskovic, O., Falanga, K., Aivazidis, K., Krstic, K., Videnovic, M., Gouveia, M. J., Castro Silva, J., Delzepich, R., Holder, L., & Enoch, C. (2019). *Enhancing student self-reflection on the situation at university. The SRT scale inventory*. <https://doi.org/10.1310/RG.2.2.23089.53600>
- O'Donovan, J., & Smyth, B. (2005, January). Trust in recommender systems. In *Proceedings of the 10th international conference on intelligent user interfaces, San Diego*, 167-174. https://www.researchgate.net/profile/Barry-Smyth/publication/221608315_Trust_in_recommender_systems/links/0fcfd50f3fa6b4ec86000000/Trust-in-recommender-systems.pdf
- OECD (2018). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing. <https://doi.org/10.1787/eag-2018-en>
- OECD (2020). *Education at a Glance 2020: OECD Indicators*. Paris: OECD Publishing. <https://doi.org/10.1787/69096873-en>
- *O'Neill, L. D., Wallstedt, B., Eika, B., & Hartvigsen, J. (2011). Factors associated with dropout in medical education: A literature review. *Medical Education*, 45(5), 440-454. <https://doi.org/10.1111/j.13652923.2010.03898.x>
- Oppedisano, V. (2009). *Open University Admission Policies and Drop Out Rates in Europe*. <https://ideas.repec.org/p/ucd/wpaper/200944.html>
- O'Regan, L., Brown, M., Harding, N., McDermott, G., & Ryan, S. (2016). *Technology-enabled feedback in the first year: A synthesis of the literature*. <https://research.thea.ie/handle/20.500.12065/2954>

References

- Oreopoulos, P., & Petronijevic, U. (2013). Making college worth it: A review of the returns to higher education. *The Future of Children*, 23(1), 41–66. <https://doi.org/10.1353/foc.2013.0001>
- Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26(1), 136-144. <https://doi.org/10.7334/psicothema2013.259>
- **Pagan, R., & Edwards-Wilson, R. (2002). A mentoring program for remedial students. *Journal of College Student Retention: Research, Theory & Practice*, 4(3), 207-226. <https://doi.org/10.2190/UFGM-8014-894V-CXFL>
- Park, J. H., & Choi, H. J. (2009). Factors influencing adult learners' decision to drop out or persist in online learning. *Journal of Educational Technology & Society*, 12(4), 207-217. <http://www.jstor.org/stable/jeductechsoci.12.4.207>
- *Pascarella, E. T. (1980). Student-faculty informal contact and college outcomes. *Review of Educational Research*, 50(4), 545-595. <https://doi.org/10.2307/1170295>
- Patterson, B., & McFadden, C. (2009). Attrition in online and campus degree programs. *Online Journal of Distance Learning Administration*, 12(2). <http://www.westga.edu/wdistance/ojdl/summer122/patterson112.html>
- **Patterson, D. A., Waya, S. W., Ahuna, K. H., Tinnesz, C. G., & Vanzile-Tamsen, C. (2014). Using self-regulated learning methods to increase Native American College retention. *Journal of College Student Retention: Research, Theory & Practice*, 16(2), 219-237. <http://dx.doi.org.10.2190/CS.16.2.d>
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, 53(3), 801-813. <https://doi.org/10.1177/0013164493053003024>
- Pinxten, M., Van Soom, C., Peeters, C., De Laet, T., & Langie, G. (2019). At-risk at the gate: prediction of study success of first-year science and engineering students in an open-admission university in Flanders—Any incremental validity of study strategies? *European Journal of Psychology of Education*, 34(1), 45-66. <https://doi.org/10.1007/s10212-017-0361-x>
- Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1), 37-82. <http://www.jstor.org/stable/2586947>
- R Core Team. (2018). *A language and environment for statistical computing* [Computer software]. <https://cran.r-project.org/>
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145-159. <https://doi.org/10.1007/s11336-008-9102-z>
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353. <https://doi.org/10.1037/a0026838>
- Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action Evaluation Framework: A review of evidence-based learning analytics interventions at the Open University UK. *Journal of Interactive Media in Education*, 1(2), 1-11. <http://doi.org/10.5334/jime.394>
- *Riggert, S. C., Boyle, M., & Petrosko, J. M. (2006). Student employment and higher education: Empiricism and contradiction. *Review of Educational Research*, 76(1), 63-92. <https://doi.org/10.3102/00346543076001063>
- *Robbins, S. B., Lauver, K. Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261. <https://doi.org/10.1037/0033-2909.130.2.261>
- Robinson, D. A. (1996). Orientation programs: A foundation for student learning and success. *New Directions for Student Services*, 75, 55-68. <https://doi.org/10.1002/ss.37119967507>
- Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education*, 6(1), 1-16. [https://doi.org/10.1016/S1096-7516\(02\)00158-6](https://doi.org/10.1016/S1096-7516(02)00158-6)
- Royal, K. D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice*, 8, 567. <https://doi.org/10.2147/AMEP.S139492>
- **Ruthig, J. C., Perry, R. P., Hall, N. C., & Hladkyj, S. (2004). Optimism and attributional retraining: Longitudinal effects on academic achievement, test anxiety, and voluntary course withdrawal in college students. *Journal of Applied Social Psychology*, 34(4), 709-730. <https://doi.org/10.1111/j.1559-1816.2004.tb02566.x>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68. <https://doi.org/10.1037/0003-066X.55.1.68>
- Sackett, P. R., Walmsley, P. T., Koch, A. J., Beatty, A. S., & Kuncel, N. R. (2016). Predictor content matters for knowledge testing: Evidence supporting content validation. *Human Performance*, 29(1), 54-71. <https://doi.org/10.1080/08959285.2015.1120307>

- Sage, A. J., Cervato, C., Genschel, U., & Ogilvie, C. A. (2018). Combining academics and social engagement: A major-specific early alert method to counter student attrition in science, technology, engineering, and Mathematics. *Journal of College Student Retention: Research, Theory & Practice*, 22(4), 611-626. <https://doi.org/10.1177/1521025118780502>
- **Salinitri, G. (2005). The effects of formal mentoring on the retention rates for first-year, low achieving students. *Canadian Journal of Education/Revue canadienne de l'education*, 28, 853-873. <https://doi.org/10.2307/4126458>
- Sandoval, W. A. (2013). Educational design research in the 21st century. In R. Luckin, S. Puntambekar, P. Goodyear, B. L. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of Design in Educational Technology* (pp. 388-396). Routledge, Taylor and Francis Group Ltd.
- Sargeant, J. M., Mann, K. V., Van der Vleuten, C. P., & Metsemakers, J. F. (2009). Reflection: a link between receiving and using assessment feedback. *Advances in Health Sciences Education*, 14(3), 399-410. <https://doi.org/10.1007/s10459-008-9124-4>
- Schlusmans, K., & Winkels, J. (2017). Hoe organiseer je hoger onderwijs op afstand? De ervaringen van de Open Universiteit. [How to organize higher distance education? Experiences of the Open University]. *Thema Hoger Onderwijs*, 5, 78-82.
- Schreurs, S. (2020). *Selection for Medical School: the quest for validity*. [Doctoral dissertation, Maastricht University].
- Schreurs, S., Cleland, J., Muijtjens, A. M., Oude Egbrink, M. G., & Cleutjens, K. (2018). Does selection pay off? A cost-benefit comparison of medical school selection and lottery systems. *Medical education*, 52(12), 1240-1248. <https://doi.org/10.1111/medu.13698>
- Seaman, J. E., Allen, I. E., & Seaman, J. (2018). *Grade Increase: Tracking Distance Education in the United States*. <https://www.onlinelearningsurvey.com/highered.html>
- Sedrakyan, G., Mannens, E., & Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50, 19-38. <https://doi.org/10.1016/j.jvlc.2018.11.002>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Simpson, O. (2006). Predicting student success in open and distance learning. *Open Learning*, 21(2), 125-138. <https://doi.org/10.1080/02680510600713110>
- **Simpson, O. (2008). Motivating learners in open and distance learning: Do we need a new theory of learner support? *Open Learning*, 23(3), 159-170. <https://doi.org/10.1080/02680510802419979>
- Simpson, O. (2010). 22%-can we do better? *The CWP Retention Literature Review*, Open University, Milton Keynes.
- Simpson, O. (2013). Student retention in distance education: Are we failing our students? *Open Learning: The Journal of Open, Distance and e-Learning*, 28(2), 105-119. <https://doi.org/10.1080/02680513.2013.847363>
- Sjöberg, L. (2015). Correction for faking in self-report personality tests. *Scandinavian Journal of Psychology*, 56(5), 582-591. <https://doi.org/10.1111/sjop.12231>
- **Sneyers, E., & De Witte, K. (2017). The effect of an academic dismissal policy on dropout, graduation rates and student satisfaction. Evidence from the Netherlands. *Studies in Higher Education*, 42(2), 354-389. <https://doi.org/10.1080/03075079.2015.1049143>
- **Stegers-Jager, K. M., Cohen-Schotanus, J., Splinter, T. A., & Themmen, A. P. (2011). Academic dismissal policy for medical students: Effect on study progress and help-seeking behaviour. *Medical Education*, 45(10), 987-994. <https://doi.org/10.1111/j.1365-2923.2011.04004.x>
- Sommer, R. (1968). The Hawthorne dogma. *Psychological Bulletin*, 70(6), 592-595. <https://doi.org/10.1037/h0026728>
- Soppe, K. F. B., Wubbels, T., Lepaal, H. J., Klugkist, I., & Wijngaards-de Meij, L. D. N. V. (2019). Do they match? Prospective students' experiences with choosing university programmes. *European Journal of Higher Education*, 9(4), 359-376. <https://doi.org/10.1080/21568235.2019.1650088>
- Stinebrickner, R., & Stinebrickner, T. R. (2014). A major in science? Initial beliefs and final outcomes for college major and dropout. *Review of Economic Studies*, 81(1), 426-472. <https://doi.org/10.1093/restud/rdt025>
- St-Onge, C., Young, M., Eva, K.W., & Hodges, B. (2017). Validity: one word with a plurality of meanings. *Advances in Health Sciences Education*, 22(4), 853-867. <https://doi.org/10.1007/s10459-016-9716-3>
- The Jamovi Project. (2019). *Jamovi*. (Version 1.0) [Computer Software]. <https://www.jamovi.org>
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125. <https://doi.org/10.3102/00346543045001089>
- Tinto, V. (1999) Taking student success seriously: Rethinking the first year of college. *NACADA Journal* 19(2), 5-9. <https://doi.org/10.12930/0271-9517-19.2.5>

References

- Tinto, V. (2012). *Completing college: Rethinking institutional action*. University of Chicago Press.
- *Trapmann, S., Hell, B., Hirn, J. O. W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Zeitschrift für Psychologie/Journal of Psychology*, 215(2), 132-151. <https://doi.org/10.1027/0044-3409.215.2.132>
- Trochim, W., & Donnelly, J. (2006). *The research methods knowledge base* (3rd ed). Atomic Dog Publishing.
- Ucar, H., & Bozkurt, A. (2019). Using motivation, volition, and performance model to overcome online procrastination. In T. Erdogan, & S. Özer (Eds.), *Current Academic Studies in Social Sciences*, (pp. 105-112). <https://doi.org/10.6084/m9.figshare.9789026.v1>
- Van den Akker, J., Bannan, B., Kelly, A. E., Nieveen, N., & Plomp, T. (2013). *Educational design research: An introduction*. Educational design research. SLO.
- Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology*, 22(5), 339-351. <https://doi.org/10.1080/0044929031000>
- Van Herpen, S. G., Meeuwisse, M., Hofman, W. A., Severiens, S. E., & Arends, L. R. (2017). Early predictors of first-year academic success at university: Pre-university effort, pre-university self-efficacy, and pre-university reasons for attending university. *Educational Research and Evaluation*, 23(1-2), 52-72. <https://doi.org/10.1080/13803611.2017.1301261>
- Van Klaveren, C., Kooiman, K., Cornelisz, I., & Meeter, M. (2019). The higher education enrollment decision: Feedback on expected study success and updating behavior. *Journal of Research on Educational Effectiveness*, 12(1), 67-89. <https://doi.org/10.1080/19345747.2018.1496501>
- *Van Rooij, E., Brouwer, J., Fokkens-Bruinsma, M., Jansen, E., Donche, V., & Noyens, D. (2018). A systematic review of factors related to first-year students' success in Dutch and Flemish higher education. *Pedagogische Studiën*, 94(5), 360-404. <http://pedagogischestudien.nl/search?identifier=644595>
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement*, 59, 197- 210. <http://dx.doi.org/10.1177/00131649921969802>
- Vossensteyn, J. J., Stensaker, B., Kottman, A., Hovdhaugen, E., Jongbloed, B., Wollscheid, S., Kaiser, F., & Cremonini, L. (2015). *Dropout and completion in higher education in Europe*. <https://op.europa.eu/en/publication-detail/-/publication/4deeebf5-0dcd-11e6-ba9a-01aa75ed71a1>
- Vulperhorst, J. P., van der Rijst, R. M., Holmegaard, H. T., & Akkerman, S. F. (2021). Unravelling why students do or do not stay committed to a programme when making a higher education choice. *Journal of Further and Higher Education*, 1-16. <https://doi.org/10.1080/0309877X.2021.1986686>
- Wachen, J., Pretlow, J., & Dixon, K. G. (2016). Building College Readiness: Exploring the Effectiveness of the UNC Academic Summer Bridge Program. *Journal of College Student Retention: Research, Theory & Practice*, 20(1), 116-138. <https://doi.org/10.1177/1521025116649739>
- **Wang, H., & Grimes, J. W. (2000). A systematic approach to assessing retention programs: Identifying critical points for meaningful interventions and validating outcomes assessment. *Journal of College Student Retention: Research, Theory & Practice*, 2(1), 59-68. <https://doi.org/10.2190/HYY4-XTBH-RFD-LU5Y>
- Warmer, J. & Kleppe, A. (2001). *Praktisch UML* [Practical UML] (2nd Ed). Addison Wesley/Pearson Education.
- Warps, J., Nooij, J., Muskens, M., Kurver, B., & Van den Broek, A. (2017). *De studiekeuzecheck: Landelijk onderzoek naar uitvoering en opbrengsten van de studiekeuzecheck*. [The study choice check: National survey on the implementation and yields of the study choice check]. ResearchNED.
- Waschull, S. B. (2005). Predicting success in online psychology courses: Self-discipline and motivation. *Teaching of Psychology*, 32(3), 190-192. https://doi.org/10.1207/s15328023top3203_11
- Wedemeyer, C. A. (2010). *Learning at the back door: Reflections on nontraditional learning in the lifespan*. Information Age Publishing.
- Willcoxson, L., Cotter, J., & Joy, S. (2011). Beyond the first-year experience: The impact on attrition of student experiences throughout undergraduate degree studies in six diverse universities. *Studies in Higher Education*, 36(3), 331-352. <https://doi.org/10.1080/03075070903581533>
- Wools, S., Eggen, T., & Sanders, P. (2010). Evaluation of validity and validation by means of the argument-based approach. *CADMO*, 8, 63-82. <https://doi.org/10.3280/CAD2010-001007>
- Wosnitzer, M., & Beltman, S. (2012). Learning and motivation in multiple contexts: The development of a heuristic framework. *European Journal of Psychology of Education*, 27(2), 177-193. <https://doi.org/10.1007/s10212-011-0081-6>
- Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 27(2), 147-170. <https://doi.org/10.1177/0265532209349465>

- Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, 18(3), 285-297. <https://doi.org/10.1080/13562517.2012.719154>
- Yigit, S., & Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study. *Revstat Statistical Journal*, 16(3), 295-313.
- Zafar, B. (2011). How do college students form expectations? *Journal of Labor Economics*, 29(2), 301-348. <https://doi.org/10.1086/658091>
- Zhang, X., Gossett, C., Simpson, J., & Davis, R. (2019). Advising students for success in higher education: An all-out effort. *Journal of College Student Retention: Research, Theory & Practice*, 21(1), 53-77. <https://doi.org/10.1177/1521025116689097>
- Zhao, X., Lynch Jr, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, 37(2), 197-206. <https://doi.org/10.1086/651257>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123-133. <https://doi.org/10.1007/s11336-003-0974-7>



Dankwoord



Dankwoord

Een goed begin is het halve werken zonder slag of stoot. Zomaar twee uitspraken die op dit promotietraject in elk geval niet van toepassing zijn. Ietwat naïef, aardig besluiteloos, en toch standvastig begon ik eind 2017 aan deze reis. De combinatie van die drie eigenschappen verklaart wellicht voor een deel waarom die eerste twee uitspraken niet van toepassing zijn. En vooral ook waarom enige sturing en steun wenselijk was gedurende dit traject. Hoe dan ook, het is wel *eind goed al goeden* dat is te danken aan een heleboel mensen, waarvoor dit hoofdstuk is bedoeld. Een kleine disclaimer vooraf: zoals de meeste mensen die in dit hoofdstuk aan bod komen al weten, ik ben alles behalve kort van stof. Om het nog ingewikkelder te maken, voor het bedanken van sommige mensen ga ik over op de Engelse taal.

First and foremost

Ik kan niet anders dan starten met het bedanken van twee belangrijke vrouwen voor mij de afgelopen 4,5 jaar. José en Kim, jullie waren het niet altijd met mij eens (en terecht!), ook niet altijd met elkaar (acceptabel), en ook niet altijd met jullie verleden zelf (interessant en ietwat verwarrend). Het kon niet uitblijven dat ik daar nu iets over zou zeggen. Nochtans kwamen we er altijd wel uit en staat buiten kijf dat ik enorm veel van jullie geleerd heb. José, jij leerde mij veel over het beargumenteren en onderbouwen van overwegingen, over onderzoek doen en daarover rapporteren. Maar ik zou de ruimte hier liever gebruiken om te benadrukken dat ik dankbaar ben voor de band die we hebben opgebouwd. Ik denk met een warm gevoel terug aan onze gesprekken, inhoudelijk, maar ook over allerlei levenskwesties en willekeurige onderwerpen. De deur staat altijd open en ik kom graag nog vaak bijkletsen! Kim, jij leerde mij mezelf beter kennen, oog houden voor mijn eigen welzijn (daar kwam af en toe een reep chocolade aan te pas), en onderzoekstechnisch de praktische relevantie niet vergeten. Ik kijk met een fijn gevoel terug op dat ik altijd onbevangen en soms hoog in de toeren binnen kon wandelen. Dankjewel daarvoor!

Zonder promotor natuurlijk geen promotietraject. Rob, ik kon jouw werkwijze – zeker in het begin – niet altijd begrijpen. Die open zee (of bedoelde je autonomie?) duurde mij elke minuut te lang. Maar jij had van begin af aan vertrouwen dat dit een *eind goed al goed* verhaal ging worden en sprak dat ook telkens uit. Het duurde even, maar dat heeft mij ook vertrouwen (of in jouw termen, gevoel van competentie) gegeven. Dankjewel daarvoor!

Collega's

In elke fase van dit traject hebben collega's een belangrijke rol gespeeld, zelfs nog voordat ik een voet binnen de Open Universiteit had gezet. Collega's bij Fontys Hogeschool Pedagogiek in Sittard – de psych chicks in het bijzonder – hebben mij altijd zoveel veren in mijn achterste gestoken dat ik dit avontuur aandurfde. Ik ben jullie niet vergeten, dankjewel voor de fijne tijd en jullie steun!

Ik prijs mezelf gelukkig dat ik ook tijdens het promotietraject een grote groep collega's om me heen had waar ik altijd op kon terugvallen. Te beginnen bij de collega's van de voormalige vakgroep 'FEEEL', inmiddels 'Voorwaarden voor een Leven Lang

Leren': Renate, Desiree, Celeste, Inge, Joyce, Jérôme, Kate, Petra, en Mara. Ook een dikke dankjewel naar de andere vakgroepen, die mij vaak betrokken hebben en interesse toonden in deze 'vakgroep overstijgende'-PhD. Dankjewel ook aan Marina en Daniëlle, voor jullie snelle service bij willekeurige vragen en praktische zaken, jullie zijn onmisbaar. En een speciale dankjewel aan de collega's waarmee ik mijn onderwijstaken vervulde: Femke, Olga, Corrie, Stefan, Kim en Migchiel.

Aan Renate en Mara wil ik graag nog een bijzonder woordje wijden. Renate, je was dan niet mijn promotor, maar hield altijd vinger aan de pols. Niet dat ik dan stond te springen om plotselinge onderzoek gerelateerde vragen te beantwoorden, maar je vroeg ook regelmatig hoe het met mij ging. En of ik al verloofd was (nog steeds nee), en of ik inmiddels al een hond had (ja). Je staat op speed dial als er een ring in het spel is, dankjewel voor alles! Mara, een van de weinige mensen die zo veel lak heeft aan mijn allergie voor eet- en drinkgeluiden op kantoor. Maar ook een van de weinige mensen die het kan hebben als ik daar een net te onaardige opmerking over maak. Ik vind je te gek, mijn forever nummer 55 (voor alle andere mensen, dit is en blijft een inside joke, beter voor iedereen). Dankjewel voor alle chocoladerepen die we gedeeld hebben, de avonden met wijn en entertainment (iets met een kruk en iemand die viel). Ik ben blij dat jij mijn paranimf wilde zijn!

Mijn promotieonderzoek kwam voort uit een al lopend project, waar een aantal mensen al goed de toon voor mij heeft gezet. Dank aan alle betrokkenen van het yOUpractice project. De zelftest in dit proefschrift was ook zeker niet tot stand gekomen zonder Hubert, Harrie, Steven en Henry. We spraken niet altijd elkaars taal (het ≠ behoort blijkbaar niet tot Python programmeertaal). Desondanks ben ik jullie dankbaar voor jullie geduld, service en alle hulp bij het bewerkstelligen van dit resultaat.

Nu volgt een waanzinnige willekeur van mensen die elkaar onderling waarschijnlijk (nog) niet kennen, maar allemaal een plekje in dit hoofdstuk verdienen. Thanks to my colleagues in the yOUng board: Daniele, Katya, Manon, en Mari. Our brainstorm sessions, events and even the table football fotoshoot were welcome distractions. Dankjewel ook aan Izaak, waarmee ik onverhoopt aan een picknick tafel belandde tijdens de ORD van 2019 in Heerlen. Fijn dat we sindsdien elkaars extra paar ogen konden zijn voor conceptversies van stukken. Ik kijk er naar uit onze nog niet uitgewerkte ideeën samen op te pakken! Nog zo'n onverwachte samenkomst was Juliët, we vormden een leftover duo bij een ICO cursus. In hindsight een dreamteam. Dank voor de talloze schrijf- en reflectiesessies. Maar ook voor de Zwitserleven wandelingen, fijne gesprekken en uiterst sarcastische humor. Dat iemand helemaal uit Hoorn dat aankan met iemand uit het zuidelijkste puntje van Limburg (inclusief accent), zegt alles. Also thanks to a bunch of people who were not as closely involved in any of my research or teaching activities, but from time to time checked in on me, sent helpful information and shared their insights. This is a non-exhaustive list but thank you, Sjef Stijnen, Jos Claessen, Ormond Simpson, and Dominique Sluijsmans.

I can only end this OU chapter by thanking the people with whom I (more or less) started it. We share a reputation at ICO, numerous long nights in strange pubs, many

inside jokes – enough to fill the Wolga with – but mostly a journey that I'll never forget. Thank you, Stefan, Selina, and Mari. I demand you make the profile picture of the next WhatsApp group one of me, so you won't forget me either. Mari, special thanks to you for being my paranympth!

Een jaar geleden maakte ik de transfer van de OU naar de Universiteit van Maastricht. Bij het departement Educational Research and Development ben ik heel warm ontvangen. Waarvoor dank aan alle ERD collega's! In het bijzonder dankjewel aan Wim, Simon, Henny, Manon, Marloes, Therese, Inken, en Alexandra, voor jullie vertrouwen, begrip en support. I am very happy to have started this new adventure together with you, Alexandra! Couldn't imagine a better partner in crime.

Familie, vrienden en de liefde

Voor (schoon)familie en vrienden was het niet altijd even duidelijk wat ik nou precies deed. En dat is ok. Voor mij was dat namelijk ook niet altijd even helder. Ik kan niet iedereen bij naam en toenaam noemen, al zou ik het willen. Iedereen die mijn familie en met name mijn schoonfamilie kent, weet dat dit tientallen pagina's zou kosten. Hoe dan ook, dank aan mijn familie: pap, mam, Lars, opa, en oma[†]. En dank aan mijn schoonfamilie: Resi, John, Silvia, Dion, Daphne, Emma en Merle. Een speciaal woord van dank aan die laatste twee. Met al dat serieuze nadenk- en schrijfwerk de afgelopen jaren waren zij een bron van relativering. Emma (10) stelt vragen over thema's waarvan men niet zou verwachten dat ze daar serieus over nadenkt en heeft zo'n goed oog voor alles wat er in haar omgeving gebeurt dat je niets voor haar kunt achterhouden. Merle (7) stelt daarentegen niet per se zoveel vragen, maar heeft oneliners en wijsheden waar de gemiddelde volwassene een puntje aan kan zuigen (zie de quote waar dit proefschrift mee begint). Emma en Merle, verander alsjeblieft niet, jullie zijn fantastisch.

Hoewel ik een deel van de mensen die ik tot nu toe heb genoemd ook onder de categorie vrienden zou kunnen noemen, zijn er toch een paar – twee paren om precies te zijn – in het bijzonder. Maikel & Marloes en Rick & Vera, dank voor de talloze borreelplanken, bankhang sessies die veranderden in heuse dance-offs, wandelingen, vakanties en dan mis ik vast nog een heleboel. Dank voor jullie luisterend oor en voor alle welkome afleiding en ontspanning. Ik word blij van jullie! Speciale dank aan Vera voor het ontwerpen van dit boekje. Je hebt het verhaal feilloos opgepikt (en ik kan wel stellen dat het niet aan de kwaliteiten van de verteller lag) en ik ben ontzettend blij met het prachtige resultaat. Je bent een topper!

Er is geen andere optie dan dat ik dit verhaal eindig met de liefde. Lieve Mickey, lieve bul (sorry voor deze exposure, ik vind dit een gegronde reden), ontelbare dank. Voor het luisteren naar allerlei overwegingen en hersenspinsels die vaak alles behalve duidelijk waren. Voor het begrijpen wanneer ik op de bank neergezet moest worden met een glaasje wijn en wat borreelnootjes. Voor mij een schop onder de kont geven als ik van alles aan het uitstellen was. Voor mij meesleuren naar buiten voor wat frisse lucht wanneer ik zelf niet wist dat ik die goed kon gebruiken. Maar vooral voor dat je mij altijd aan het lachen maakt. Keer op keer, ook als ik denk dat er niets te lachen valt. Ik hou van je.



ICO Dissertation Series



ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, Radboud University Nijmegen, University of Amsterdam, University of Antwerp, University of Ghent, KU Leuven, Université Catholique de Louvain (since 2018), University of Groningen, University of Twente, Utrecht University, Vrije Universiteit Amsterdam, and Wageningen University, and formerly Tilburg University (until 2002). The University of Groningen, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

304. Wijnen, M. (01-02-2019) *Introduction of problem-based learning at the Erasmus School of Law: Influences on study processes and outcomes*. Rotterdam: Erasmus University Rotterdam
305. Dobbelaer, M.J. (22-02-2019) *The quality and qualities of classroom observation systems*. Enschede: University of Twente
306. Van der Meulen, A.N. (28-02-2019) *Social cognition of children and young adults in context*. Amsterdam: Vrije Universiteit Amsterdam
307. Schep, M. (06-03-2019) *Guidance for guiding. Professionalization of guides in museums of art and history*. Amsterdam: University of Amsterdam
308. Jonker, H.M. (09-04-2019) *Teachers' perceptions of the collaborative design and implementation of flexibility in a blended curriculum*. Amsterdam: University of Amsterdam
309. Wanders, F. H. K. (03-05-2019). *The contribution of schools to societal participation of young adults: The role of teachers, parents, and friends in stimulating societal interest and societal involvement during adolescence*. Amsterdam: University of Amsterdam
310. Schrijvers, M.S.T. (03-05-2019) *The story, the self, the other. Developing insight into human nature in the literature classroom*. Amsterdam: University of Amsterdam
311. Degrande, T. (08-05-2019) *To add or to multiply? An investigation of children's preference for additive or multiplicative relations*. Leuven: KU Leuven.
312. Filius, R.M. (23-05-2019) *Peer feedback to promote deep learning in online education. Unravelling the process*. Utrecht: Utrecht University
313. Woldman, N. (24-05-2019) *Competence development of temporary agency workers*. Wageningen: Wageningen University
314. Donszelman, S. (06-06-2019) *Doeltaal-leertaal didactiek, professionalisering en leereffecten*. Amsterdam: Vrije Universiteit Amsterdam
315. Van Oeveren, C.D.P. (12-06-2019) *ITHAKA gaf je de reis*. Amsterdam: Vrije Universiteit Amsterdam
316. Agricola, B.T. (21-06-2019) *Who's in control? Finding balance in student-teacher interactions*. Utrecht: Utrecht University
317. Cuyvers, K. (28-08-2019), *Unravelling medical specialists self-regulated learning in the clinical environment*. Antwerp: University of Antwerp
318. Vossen, T.E. (04-09-2019) *Research and design in STEM education*. Leiden: Leiden University
319. Van Kampen, E. (05-09-2019) *What's CLIL about bilingual education?* Leiden: Leiden University
320. Henderikx, M.A. (06-09-2019) *Mind the Gap: Unravelling learner success and behaviour in Massive Open Online Courses*. Heerlen: Open University of the Netherlands
321. Liu, M. (13-09-2019) *Exploring culture-related values in Chinese student teachers' professional self-understanding and teaching experiences*. Utrecht: Utrecht University
322. Sun, X. (13-09-2019) *Teacher-student interpersonal relationships in Chinese secondary education classrooms*. Utrecht: Utrecht University

323. Wu, Q. (02-10-2019) *Making Construct-Irrelevant Variance Relevant: Modelling item position effects and response behaviors on multiple-choice tests*. Leuven: KU Leuven
324. Jansen, R.S. (11-10-2019) *Dealing with autonomy: Self-regulated learning in open online education*. Utrecht: Utrecht University
325. Van Ginkel, S.O. (23-10-2019) *Fostering oral presentation competence in higher education*. Wageningen: Wageningen University
326. Van der Zanden, P. (05-11-2019) *First-year student success at university: Domains, predictors, and preparation in secondary education*. Nijmegen: Radboud University Nijmegen
327. De Bruijn, A.G.M. (14-11-2019) *The brain in motion: Effects of different types of physical activity on primary school children's academic achievement and brain activation*. Groningen: University of Groningen
328. Hopster-Den Otter, D. (28-11-2019) *Formative assessment design: A balancing act*. Enschede: University of Twente
329. Harmsen, R. (10-12-2019) *Let's talk about stress. Beginning secondary school teachers' stress in the context of induction programmes*. Groningen: University of Groningen
330. Post, T. (11-12-2019) *Fostering inquiry-based pedagogy in primary school: a longitudinal study into the effects of a two-year school improvement project*. Enschede: University of Twente
331. Ackermans, K. (20-12-2019) *Designing Video-Enhanced Rubrics to Master Complex Skills*. Heerlen: Open University of the Netherlands
332. Li, N (08-01-2020) *Analyzing online in-service teacher training courses in China*. Eindhoven: Eindhoven University of Technology
333. Figueroa Esquivel, F. (13-02-2020) *Early childhood multidimensional development. A rapid and non-linear roller coaster*. Groningen: University of Groningen
334. De Leeuw, R.R. (27-02-2020) *Through the eyes of the beholder. Unfolding social participation "from within" the classroom*. Groningen: University of Groningen
335. Schreurs, S. (20-03-20) *Selection for medical school. The quest for validity*. Maastricht: Maastricht University
336. Nugteren, M.L. (12-06-2020) *What do I know and where do I go? The effects of guidance on task selection*. Utrecht: Utrecht University
337. Wijns, N. (01-07-2020) *On the hunt for regularities: An investigation of children's early patterning competencies*. Leuven: KU Leuven
338. Duijzer, A.C.G. (25-08-2020) *Reasoning about graphs in primary mathematics education*. Utrecht: Utrecht University
339. Dijkema, S. (25-08-2020) *Ready for takeoff? The relation between the type of teacher training program and daily teaching practices of Dutch beginning primary school teachers*. Groningen: University of Groningen
340. Slot, E.M. (11-09-2020) *Characterizing adolescents' interest: understanding multiplicity and dynamics in persons, objects and contexts*. Utrecht: Utrecht University
341. Ping, C. (24-09-2020) *Understanding Teacher Educators' Professional Learning*. Eindhoven: Eindhoven University of Technology
342. Van Rijswijk, M.M. (16-10-2020) *Experiences of continuity and discontinuity in student teachers' development*. Utrecht: Utrecht University
343. Rovers, S.F.E. (16-09-2020) *Growing knowledge; Supporting students' self-regulation in problem-based learning* Maastricht: Maastricht University

344. Langeloo, A. (17-09-2020) *Multilingual and monolingual children in kindergarten classrooms: Exploring teacher-child interactions and engagement as learning opportunities*. Groningen: University of Groningen
345. Sekeris, E. (06-10-2020) *Unravelling computational estimation development in 5- to 7-year-olds* Leuven: KU Leuven
346. Ter Beek, M. (29-10-2020) *Supporting reading comprehension in history education: The use and usefulness of a digital learning environment* Groningen: University of Groningen
347. Peppen, L.M. (25-09-2020) *Fostering Critical Thinking. Generative processing strategies to learn to avoid bias in reasoning* Rotterdam: Erasmus University Rotterdam
348. Van Geel, K., (05-11-2020) *Lifelong learning in radiology: All eyes on visual expertise* Maastricht: Maastricht University
349. Donker, M.H. (20-11-2020) *In DEPTH: Dynamics of Emotional Processes in Teachers – An exploration of teachers' interpersonal behavior and physiological responses* Utrecht: Utrecht University
350. Janssen, E.M. (13-11-2020) *Teaching Critical Thinking in Higher Education: Avoiding, Detecting, and Explaining Bias in Reasoning* Utrecht: Utrecht University
351. Van den Broek, E.W.R. (09-10-2020) *Language awareness in foreign language education. Exploring teachers' beliefs, practices and change processes* Nijmegen: Radboud University Nijmegen
352. Kasch, J.A. (09-10-2020) *Scaling the unscalable? Interaction and support in open online education* Heerlen: Open University of the Netherlands
353. Otten, M. (30-10-2020) *Algebraic reasoning in primary school: A balancing act* Utrecht: Utrecht University
354. De Vrind, E. (25-11-2020) *The SpeakTeach method, Towards self-regulated learning of speaking skills in foreign languages in secondary schools: an adaptive and practical approach* Leiden: Leiden University
355. Tacoma, S.G. (15-11-2020) *Automated intelligent feedback in university statistics education* Utrecht: Utrecht University
356. Boonk, L.M. (04-12-2020) *Exploring, measuring, and evaluating parental involvement in vocational education and training* Heerlen: Open University of the Netherlands
357. Kickert, R. (04-12-2020) *Raising the bar: Higher education students' sensitivity to the assessment policy* Rotterdam: Erasmus University Rotterdam
358. Van der Wal, N.J. (09-12-2020) *Developing Techno-mathematical Literacies in higher technical professional education* Utrecht: Utrecht University



Open Universiteit