International Journal of Assessment Tools in Education
2022, Vol. 9, No. 2, 376–396
https://doi.org/10.21449/ijate.949160
Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate *Research Article*

Towards an Online Self-Assessment for Informed Study Decisions–A Mixed-Methods Validation Study

Laurie E. C. Delnoij[1,2*], **José P. W. Janssen**[1], **Kim J. H. Dirkx**[1,3], **Rob L. Martens**[1]

[1]Open Universiteit, Faculty of Educational Sciences, Valkenburgerweg, 177, 6419 AT, Heerlen, the Netherlands
[2]Maastricht University, School of Business and Economics, Department of Educational Research and Development, Tongersestraat 53, 6211 LM, Maastricht, the Netherlands
[3]Zuyd Hogeschool, Nieuw Eyckholt 300, 6419 DJ Heerlen, the Netherlands

ARTICLE HISTORY
Received: Sep. 09, 2021
Revised: Mar. 28, 2022
Accepted: Apr. 08, 2022

Keywords:
Study decisions,
Self-assessment,
Validity,
Mixed-methods,
Higher education.

Abstract: Informed study decisions are pivotal for student retention in higher online education. A self-assessment prior to enrolment has been proposed as a promising approach to enable informed decision-making and to build resources for retention. To determine whether such a self-assessment affects the decision-making process as intended, thorough and careful validation is a necessity. This study reports on two validity aspects that are less commonly addressed in that respect, but essential for evaluating effectiveness: response processes and consequences of (self-) testing. To map the response processes and consequences of the current self-assessment, a mixed-methods approach was used in which eight prospective students took a self-assessment in an observed think-aloud mode and were interviewed before and after that. Results show different response processes depending on the type of subtest that is taken. The results also indicate that 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.

1. INTRODUCTION

Adequate, personalized 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, 2005; 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

*CONTACT: Laurie E. C. Delnoij ✉ l.delnoij@maastrichtuniversity.nl Maastricht University, School of Business and Economics, Department of Educational Research and Development, Tongersestraat 53, 6211 LM, Maastricht, the Netherlands.

e-ISSN: 2148-7456 /© IJATE 2022

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.

1.1. Self-assessments 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 a 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; Delnoij et al., 2020b; Delnoij et al., 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., 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).

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 (American Educational Research Association [AERA] et al., 1999): 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, 1988; 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; Delnoij et al, 2020b; Delnoij et al, 2021), the present study aims to investigate *response processes* and *consequences* of a self-assessment for informed study decisions.

1.3. Process and Consequential Aspects of Validity

The process aspect of validity 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.

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., 1999; 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., 2016). 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 enroll. 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.

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?*

2. METHOD

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 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).

2.2. Research 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 enroll) 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).

2.3. Materials

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

2.3.1. *Self-assessment prototype*

The SA prototype, illustrated in Figure 1, consists of four constituent tests, completion of which results in a score and related feedback per subtest. The subtests measure numerical skills, 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 color 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 1, C-E).

2.3.2. *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.

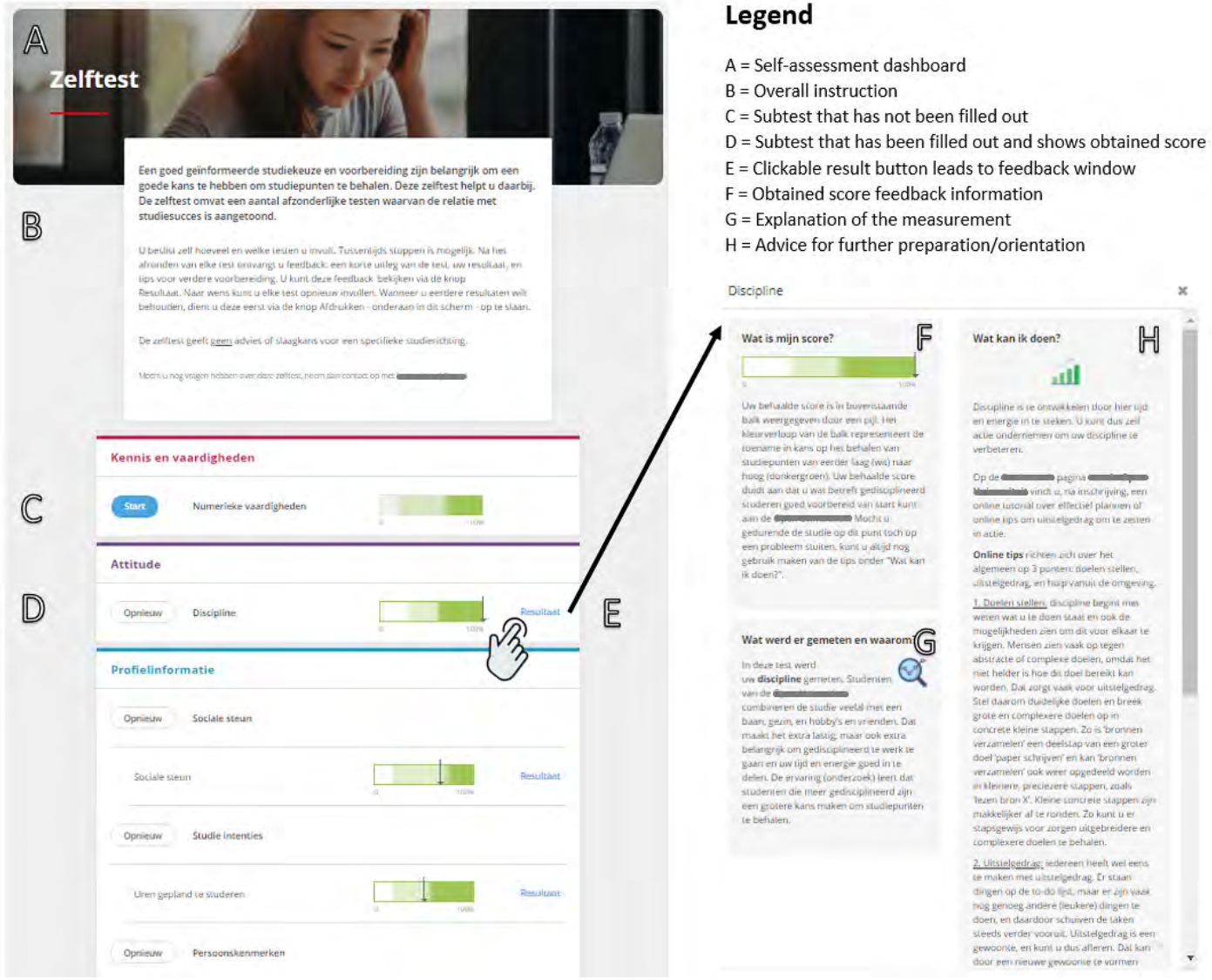
2.3.3. *Semi-structured interview protocol*

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 enroll) to 10 (certain to enroll) both prior to and after taking the SA.

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).

Figure 1. *Prototypical Self-assessment.*



2.5. Procedure

2.5.1. *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.

2.5.2. *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, 1990; 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).

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 analyzed using qualitative content analysis.

2.6.1. *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 2 or more categories, and are descriptive in name. As an example, "Process threat" and "Process evidence" were grouped together as "Process aspect of validity".

3. RESULT

3.1. Expectations of SA Impact (RQ1)

Table 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 P†). 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".

Table 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 order[1] obtained score[2] and feedback consultation[3] 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

[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

† All participants were given an anonymous identifier, obtained via Randomwordgenerator©.

3.2. Obtained Scores (RQ2)

A summary of the obtained subtests scores is provided in Table 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.

3.3. Test Taking Behavior and Feedback Consultation (RQ3)

A summary of the number and order of subtests taken and feedback consultation is provided in Table 1.

3.3.1. *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).

3.3.2. *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).

3.3.3. *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).

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.

3.4.1. *Numerical skills*

3.4.1.1. 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.

3.4.1.2. 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.

3.4.2. *Discipline*

3.4.2.1. 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.

3.4.2.2. 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.

3.4.3. *Social support*

3.4.3.1. 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)

3.4.3.2. 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.

3.4.4. *Hours planned to study*

3.4.4.1. 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).

3.4.4.2. 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.

3.4.5. *Overall*

3.4.5.1. 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 minimize this effect, as we do not want to discourage respondents unnecessarily.

3.4.5.2. 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.

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)

3.6. Study Choice Certainty (RQ7)

A summary of participants' study choice certainty is provided in Table 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 enroll 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 thoughtvery 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)

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.

Please present the findings/results in this section. This section should give significant results obtained from the study clearly and concisely. Please present the findings/results in this section. This section should give significant results obtained from the study clearly and concisely.

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. (2020) 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.

4.1. Implications for the SA, Theory and Practice

4.1.1. *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.

4.1.2. *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).

4.1.3. *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; Robbinson et al., 1996; Sage et al., 2018; Wachen et al., 2016).

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., 2012 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.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. Ethical approval was obtained by the Ethics Committee of the Open University (code: U20200563).

Authorship Contribution Statement

Laurie E. C. Delnoij: Investigation, Resources, Visualization, Formal Analysis, and Writing-original draft and revisions. **José P. W. Janssen**: Methodology, Formal Analysis, Supervision, Validation, and Writing-revision. **Kim J. H. Dirkx**: Methodology, Supervision, Validation, and Writing-revision. **Rob L. Martens**: Supervision, Validation, and Writing-revision.

Orcid

Laurie E. C. Delnoij https://orcid.org/0000-0001-6363-5714
José P. W. Janssen https://orcid.org/0000-0002-5104-7648
Kim J. H. Dirkx https://orcid.org/0000-0001-8014-0916
Rob L. Martens https://orcid.org/0000-0001-7193-8125

REFERENCES

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.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* American Psychological Association.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychologi cal Review*, *84*(2), 191-215.

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

Borghans, L., Golsteyn, B., & Stenberg, A. (2015). Does expert advice improve educational choice? *PLoS One*, *10*(12), Article e0145378. https://doi.org/10.1371/journal.pone.0145378

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*, 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

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), 68-82. https://doi.org/10.26522/brocked.v12I2.38

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, A.D. (2006). The coming age of research on test-taking strategies. *Language Assessment Quarterly*, *3*(4), 307-331. https://doi.org/10.1080/15434300701333129

Cook, 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

Creswell, J.W. (2014). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Edinburgh Gate: Pearson Education Limited.

Delnoij, L.E.C., Dirkx, K.J.H., Janssen, J.P.W., & Martens, R.L. (2020a). 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., Gijselaers, 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/feduc.2021.668922

Delnoij, L.E.C., Janssen, J.P.W., Dirkx, K.J.H., Martens, R.L. (2020b) *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.

Downing, S.M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education*, *37*(9), 830-837. https://doi.org/10.1046/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

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

Fonteyne, L., & Duyck, W. (2015). Vraag het aan SIMON! [Ask SIMON!]. *Thema Hoger Onderwijs*, *2*, 56-60.

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

Guba, E.G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, *29*(2), 75-91.

Hartnett, M. (2016). The importance of motivation in online learning. In Harnett, M. (Ed.) *Motivation in online education* (pp. 5-32). Springer. https://doi.org/10.1007/978-981-10-0700-2_2

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*. https://doi.org/10.1016/j.iheduc.2020.100758

Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *American Journal of Occupational Therapy*, *45*(3), 214-222.

Kubinger, K.D., Frebort, M., & Müller, C. (2012). Self-assessment im rahmen der studientberatung: Möglichkeiten und Grenzen. In: Kubinger, K.D., Frebort, M., Khorramdel, L., & Weitensfelder, L. (eds). *Self-Assessment: Theorie und Konzepte*, 9-24. [Self-Assessment: Theory and Concepts] Lengerich: Pabst Science Publishers.

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

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

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

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

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*. https://doi.org/10.28945/4182

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., Ilic, I.S., Almeida, L.S., Stamovlasis, D., Do Céu Taveira, M., Toskoviç, O., Falanga, K., Aivazidis, C., Krstic, K., Videnovic, M., Gouveia, M., Castro Silva, J., Delzepich, R., Holder, L., & Clinton, E. (2019). Enhancing student self-reflection on the situation at university. The SRT scale inventory.

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

Niessen, A.S.M., & Meijer, R.R. (2017). Voorspellen in het kader van de studiekeuzecheck: Tijd voor verbetering [Predicting in the context of study decisions: time for improvement]. *Onderzoek van Onderwijs*, *46*, 5-7.

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), e0198746. https://doi.org/10.1371/journal.pone/0198746

O'Regan, L., Brown, M., Harding, N., McDermott, G., & Ryan, S. (2016) *Technology-enabled feedback in the first year: A synthesis of the literature*. http://y1feedback.ie/wpcontent/uploads/2016/04/SynthesisoftheLiterature2016.pdf

Padilla, J.L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *26*(1), 136-144.

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

Royal, K.D. (2017). Four tenets of modern validity theory for medical education assessment and evaluation. *Advances in Medical Education and Practice*, *8*, 567-570. https://doi.org/10.2147/AMEP.S139492

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

Shenton, A.K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, *22*(2), 63-75.

Sommer, R. (1968). The Hawthorne dogma. *Psychological Bulletin*, *70*(6, Pt.1), 592. https://doi.org/10.1037/h0026728

Soppe, K.F.B., Wubbels, T., Leplaa, 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

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*, 853-867. https://doi.org/10.1007/s10459-016-9716-3

Tinto, V. (2005) Taking student success seriously: Rethinking the first year of college. IN: Taking student success seriously: Rethinking the first year of college. *Paper Presented at the Ninth Annual Intersession Academic Affairs Forum, California State University, Fullerton.*

Van den Akker, J., Bannan, B., Kelly, A.E., Nieveen, N., & Plomp, T (2013). *Educational design research: An introduction. Educational design research.* Enschede.

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

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 H., Kottmann, A., Jongbloed, B.W.A., Kaiser, F., Cremonini, L, Stensaker, B., Hovdhaugen, E., & Wollscheid, S. (2015). Dropout and completion in higher education in Europe: Main report. https://research.utwente.nl/en/publications/dropout-and-completion-in-higher-education-in-europe-main-report.

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

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