Reliability and validity of new online selection tests for midwifery students

Oana R. Groene [a,*], Mirjana Knorr [a], Daniela Vogel [b], Carmen Hild [b], Wolfgang Hampe [a]

[a] Institute of Biochemistry and Molecular Cell Biology, Center for Experimental Medicine, University Medical Center Hamburg-Eppendorf, 52 Martinistraße, 20246 Hamburg, Germany
[b] Academy for Education and Career, University Medical Center Hamburg-Eppendorf, 52 Martinistraße, 20246 Hamburg, Germany

ARTICLE INFO

ABSTRACT

Background and objective: Newly created midwifery bachelor programmes attract an increasing number of applicants. Hence, educators have moved away from traditional non-standardised selection methods and started developing new tools for the admission of students. Researchers have expressed concerns around the validity of such selection measures and their ability to identify the appropriate candidates. The objective of this study was to develop and implement selection procedures for a new midwifery Bachelor of Science degree in Hamburg, Germany, and assess their psychometric properties.

Design: This is a mixed-methods longitudinal study on the reliability, validity and acceptability of newly developed selection criteria.

Setting: The newly established midwifery bachelor of science in Hamburg, Germany.

Participants: Upon completion of their online application, all midwifery applicants were invited by email to participate in our research project and thereby informed of its purpose and voluntary nature of their participation. The total number of candidates who took the test was 366, of which 309 agreed to take part in this study. Sixty-five applicants were eventually admitted to the midwifery Bachelor programme, of which 59 were included in the study.

Measurements and findings: We developed two tests for cognitive ability (HAM-Mid I and II) and adapted one test for personal and professional characteristics (Casper – Computer-based Assessment for Sampling Personal Characteristics). Due to the Covid pandemic, we implemented the tests online instead of on site as initially planned. HAM-Mid I had the lowest (Cronbach's Alpha = 0.38) whereas Casper had the highest internal consistency (Cronbach's Alpha = 0.77) of all three tests. Age, previous academic achievement and native tongue were significantly associated with applicants' performance on HAM-Mid II and Casper admission tests. HAM-Mid II was associated with students' performance in the first year midwifery exam ($r = 0.31$, $p < 0.05$). Overall evaluation of HAM-Mid II and Casper was good while HAM-Mid I received a more negative feedback.

Key conclusions: Evidence on the reliability and predictive validity of the newly developed tests suggests that the applied admission criteria are appropriate for an objective selection of bachelor midwifery students.

Implications for practice: The interests of university and clinical stakeholders of a midwifery bachelor degree can be represented through the development of tests for the assessment of cognitive ability and personal and professional characteristics.

Introduction

The midwifery profession has undergone enormous changes in the past two decades. Driven by the need to strengthen the quality of healthcare delivery (Ventura et al., 2015) and societal requirements to diversify the healthcare force (Effland and Hays, 2018), midwifery education has consequently adopted reforms that included the switch from vocational training to an academic degree (Bovermann, 2020; Farley and Carr, 2003; Hermansson and Mårtensson, 2013). In some European countries like Germany, the implementation of international regulations such as the EU

---

* Corresponding author.
E-mail address: o.groene@uke.de (O.R. Groene).

directive 2013/55/EU has taken longer, which meant that until 2019 the majority of German midwives still trained in vocational training colleges (Plappert et al., 2019; Bovermann, 2020). These changes created new opportunities for stakeholders to re-examine the positioning of midwives within the healthcare system and the requirements for midwifery students' recruitment and selection. Professional associations (International Confederation of Midwives, 2AD) as well as governmental institutions (World Health Organization, 2017) define the role of today's midwives and underline the importance of required competencies described as a combination of knowledge and skills. Future midwives are required to apply evidence-based knowledge in practice, work independently, assume responsibility for their practice, and effectively communicate with women and their families. Because of the diversity of required competencies, assessment strategies in midwifery education include a variety of testing methods. Beside written exams focused on acquirement of knowledge, Objective Structured Clinical Examinations and mini clinical examination exercises are being widely used (Smith et al., 2012; Sweet et al., 2013; Yigzaw et al., 2015; Malakooti et al., 2018; Hakimi et al., 2021). Testing procedures include a series of short standardised interactions between students and simulated patients which assess students' ability to transfer knowledge in realistic clinical settings.

Guided by professional and academic standards, selection procedures for healthcare education programmes such as nursing generally consider a balance between cognitive skills necessary for the acquirement and application of knowledge and professional characteristics (Zamanzadeh et al., 2020) including social skills required for a good communication with patients or women in the case of midwifery (Pitt et al., 2014; Waugh et al., 2014; World Health Organization, 2016). Measures of cognitive attributes such as secondary school grade point average (GPA) and knowledge tests are largely used for the admission of health professions students because of their ability to differentiate between candidates and to predict their academic performance (Ferguson et al., 2002). However, such cognitive measures cannot predict professional behaviour (Siu and Reiter, 2009), therefore they are often complemented by the assessment of personal and professional characteristics (Dore et al., 2017). A particular focus on such characteristics is emphasised by midwifery professional groups, which led to the use of standardised interviews such as the multiple mini interviews (MMIs) for the selection of future midwives (Waugh et al., 2014; Callwood et al., 2019). However, in spite of their good reliability and validity (Callwood et al., 2018; Callwood et al., 2020), MMIs require many resources (Hissbach et al., 2014) and are therefore often administered to smaller, pre-selected candidate samples. An alternative to MMIs is constituted by the written "Situational Judgement Test" (SJT), that presents candidates with several short scenario descriptions followed by closed or open-ended questions (Patterson et al., 2016a). Online versions of SJTs exist whereby candidates need to assess a situation based on a video or written scenario (Dore et al., 2017). Such procedures were shown to predict personal and professional characteristics of medical students in national examination scores (Dore et al., 2017; Webster et al., 2020).

In spite of the fact that midwifery education attracts highly motivated students (Green and Baird, 2009), alarming attrition rates have raised concerns amongst educators (Carolan and Kruger, 2011; Hughes, 2013). Critics claim that universities choose selection criteria that fit the academic requirements of their institutions and not the professional expectations needed for good healthcare quality (Rodgers et al., 2013). Therefore, researchers highlight the need to gather evidence about the validity and reliability of selection procedures for nurses and midwives (Taylor et al., 2014). While evidence exists on the validity of cognitive tests for nursing (Zamanzadeh et al., 2020) or medical students (Patterson et al., 2016b), no studies have been published on the validity and psy-

chometric properties of such tests for the selection of midwifery students. At the same time, although there is some evidence on MMIs' ability to predict midwifery students' academic performance (Callwood et al., 2020), there is no research exploring the validity of other methods of assessing personal and professional characteristics such as online SJTs for midwifery students. To our knowledge, the current study is therefore the first one to present validity and reliability evidence for a combination of newly developed assessment methods of cognitive ability and personal and professional characteristics, specifically applied to midwifery programmes. Additionally, we investigated if our assumptions and previous study findings in medical education related to the potential effect of socio-demographic characteristics on admission also apply to midwifery candidates. We assumed that older age and experience could positively influence applicants' performance in the assessment of personal and professional characteristics in that they would enable applicants to generate richer arguments and multifaceted perspectives. On the other hand, in light of previous research (Puddey and Mercer, 2013) we assumed that younger age might have a positive impact on cognitive test performance because younger candidates are often closer to the school and testing environment. Finally, we were aware of evidence showing that the first language (Puddey and Mercer, 2013) and parents' academic background constitute success factors for admission (Meyer et al., 2019) with native speakers and students who come from academic households having an advantage over their peers.

The goal of our study was to develop and implement selection procedures for a new midwifery Bachelor of Science degree in Hamburg, Germany, and assess their psychometric properties. We had three main research questions. Are the newly developed tests:

(1) **reliable**? i.e. Are the test scores internally consistent?
(2) **valid?** i.e. How do test scores relate to applicants' sociodemographic characteristics, to other admission criteria and to early academic performance?
(3) **accepted** by applicants?

Methods

Study setting and participants

The selection procedure in our study took place in May 2020 and was triggered by the launch of the midwifery Bachelor of Science degree in Hamburg, Germany. The programme was conceived as a "dual" academic degree, which translates into students spending half of their time learning theory and the other half in clinical placements. The newly developed education programme benefited from the cooperation between six institutions: two university centres and four maternity hospitals in the Hamburg region which host and mentor students during clinical placements. All six cooperating institutions were involved in determining the criteria for students' selection, which included measures of cognitive ability and personal and professional characteristics. Upon completion of their online application, all midwifery applicants were invited by email to participate in our research project and thereby informed of its purpose and voluntary nature of their participation.

HAM-Mid

HAM-Mid is a two-part multiple-choice test designed to measure cognitive skills needed for the midwifery bachelor degree. Representatives of cooperating institutions developed a catalogue of learning objectives that addressed biology, social sciences, midwifery, and arithmetic problem solving (Appendix 1), which was made available online to all candidates on a registration platform. Based on the catalogue, questions were developed through a review process which ensured the appropriateness of test questions.

The HAM-Mid consisted of two subtests: HAM-Mid I and II. The two subtests intend to measure aspects of general intelligence: fluid intelligence, i.e. the ability to think logically and solve problems, and crystalline intelligence, i.e. the ability to apply previously accumulated knowledge (Cattell, 1967).

HAM-Mid I was a speed test measuring fluid intelligence, which included 15 arithmetic problem solving questions; the subtest had a time limit of 10 min and taking notes was not allowed. HAM-Mid I had been previously developed and tested on a sample of medicine applicants. HAM-Mid II was conceived as a knowledge test measuring crystalline intelligence with a time limit of one hour. It contained 40 questions related to midwifery tasks and responsibilities, social sciences and biology. Some of the questions were related to a 30 min lecture on midwifery subjects (e.g. normal birth) developed for the purpose of selection and provided online one week before the test.

The HAM-Mid test was initially planned as a paper/pencil test. In order to comply with social distancing rules during the Covid-19 pandemic we decided to conduct the HAM-Mid online. Under these circumstances we could not supervise the applicants sufficiently to prevent taking notes during the HAM-Mid I.

Casper

We originally planned to use a Multiple Mini-Interview for the assessment of personal and professional characteristics. Due to the Covid-19 pandemic we decided to use the online situational judgement test Casper (Dore et al., 2017) which has been widely used in Canada, USA and Australia for health programmes selection including midwifery (https://www.mtroyal.ca/Admission/Casper-information.htm). The test has an approximate duration of one hour and a half. Candidates had five minutes to answer three questions on eight video and four written scenarios that assessed ten personal and professional characteristics: collaboration, communication, empathy, equity, ethics, motivation, problem solving, professionalism, resilience and self-awareness (Dore et al., 2009, 2017). At the time of its administration, Casper had never been used in Germany. Because of the limited time available, we used English videos with German subtitles and offered applicants the possibility to see the video twice; however, candidates had to write their answers in German. For each of the twelve scenarios, candidates' answers were assessed by a different rater, therefore their score reflected independent assessments of personal and professional characteristics. Raters were mainly midwives, psychologists and other employees of the midwifery programme in Hamburg region. Prior to the Casper test, raters received a two hour online training that included a general introduction to Casper constructs and rating procedures as well as a scenario-specific training. The Casper developer, Altus Assessment, administered the test, checked for applicants' identity, video-supervised the applicants during the examination, and offered online support using a chat function. Immediately after the Casper test, applicants were transferred to an e-learning platform where they took the HAM-Mid test.

Weighting of selection criteria

Initially we intended to select the participants according to a ranking list based on: GPA (30 points), HAM-Mid I (6 points), HAM-Mid II (24 points), and Casper (30 points). Due to a technical error, HAM-Mid II was interrupted, had to be restarted and candidates consequently had different time limits to finish the test. Because of legal considerations, we therefore excluded HAM-Mid II from the ranking. Nevertheless, the distortion of HAM-Mid II data was small and we therefore included it in this study.

Evaluation survey and socio-demographic characteristics

After the completion of the online tests, midwifery candidates were invited to participate in an evaluation survey that aimed at assessing acceptability and feasibility of tests. The survey contained 49 questions in total (23 related to Casper and 26 related to HAM-Mid). Candidates were asked to rate their experience in using the online platforms, the difficulty, readability and fairness of the different tests. Open-ended questions facilitated candidates' perspectives on the online test experience and improvement strategies. Socio-demographic data were collected concerning participants' age, previous academic achievement (existing bachelor degree), healthcare experiences (at least six months), parents' academic background and native language (German versus other languages).

Study success: outcome criteria used for predictive validity

Admitted students took two exams testing students' cognitive ability in the first semester:

(1) Science project: a thesis written following a module on scientific practice, the content, scope and grading of which depended on the subject chosen by the student.
(2) Midwifery exam: an oral exam on midwifery subjects.

The exams were rated according to the German grading system with marks from 1 (excellent) to 6 (insufficient). In order to facilitate the interpretation of the statistical analysis results, the marks were inverted resulting in 6 (excellent) to 1 (insufficient).

Due to the Covid-19 pandemic, no exams for the assessment of social competencies took place.

Data collection and data analysis

The data on applicants' admission tests were collected in May 2020 while data on admitted students' academic performance in the first semester were collected in February 2021. We analysed data with IBM SPSS Statistics for Windows 26. We used descriptive statistics to assess the distribution of test scores, test difficulty and candidates' socio-demographic characteristics. Pearson correlation coefficients were calculated to investigate how the selection and outcome criteria related to each other. We report on Spearman Rho values to describe the correlation between outcome criteria and admission tests as the distribution of the 1st semester marks was not normal (non-parametric analysis). Reliability was determined for each test by analysis of internal consistency (Cronbach's Alpha). Stepwise multiple regression analyses using a backwards method were performed to investigate whether socio-demographic characteristics can predict admission test scores. Finally, we used a thematic analysis to describe candidates' perceptions of the test experience and appropriateness.

Results

Socio-demographic characteristics of applying and admitted candidates

Four hundred and sixty nine midwifery applicants initially registered for the midwifery programme. Three hundred and sixty six decided to take the tests, of which 309 (83.6%) agreed to take part in this study. Sixty-five applicants were admitted, of which 59 (90.8%) were included in the study. All admitted and 99% of applying students were of female gender (Table 1). Approximately a third of admitted students already had a university degree while the same proportion of students also had a minimum six months experience in the healthcare field. Only 7.4% of applicants and 3.3% of successful candidates were non-German native speakers.

Table 1
Socio-demographic characteristics of applying and admitted candidates.

Variable	Applicants (N = 309)	Admitted students (N = 59)
Female gender	306 (99%)	59 (100%)
Age (Mean, SD)	23.7 (5.5)	24.4 (5.1)
Existing bachelor degree	57 (18.4%)	19 (31.7%)
German native speakers	286 (92.6%)	58 (96.7%)
Minimum 6 months of healthcare experience	98 (31.7%)	19 (31.7%)
At least one parent with university degree	146 (47.2%)	32 (54.2%)

Table 2
Admission tests: internal consistency and correlations.

Test	Mean (SD)	Cronbach's alpha (95% CI)	Correlation to admission criteria (Pearson correlation coefficient) N = 309		Correlation to study success (Spearman correlation coefficient) N = 59	
			HAM-Mid I	HAM-Mid II	Science project	Midwifery exam
HAM-Mid I	2.3 (1.0)	0.55 (0.48 - 0.62)			0.16 (0.220)	−0.21 (0.107)
HAM-Mid II	16.9 (2.9)	0.68 (0.63 - 0.73)	0.29** (< 0.001)		−0.11 (0.402)	0.31* (0.015)
Casper	15.3 (5.0)	0.72 (0.68 - 0.77)	0.19* (0.001)	0.32** (< 0.001)	0.09 (0.479)	0.13 (0.328)

Table 3
Multiple regressions predicting admission tests based on socio-demographic characteristics.

Outcome variable: HAM-Mid I	Regression B coefficient	P	F	df	R^2	R	p
Initial model			1.065	5–291	0.018	0.134	0.380
Age	−0.007	0.602					
Native tongue	0.310	0.208					
Existing bachelor degree	0.228	0.201					
Healthcare experience	−0.062	0.620					
One parent with university degree	0.148	0.204					
Outcome variable: HAM-Mid II	Regression B coefficient	P	F	df	R^2	R	p
Initial model			6.075	5–291	0.095	0.307	<0.001
Age	0.217	<0.001					
Native tongue	4.408	<0.001					
Existing bachelor degree	0.102	0.904					
Healthcare experience	−0.384	0.514					
One parent with university degree	0.770	0.163					
Final model			13.985	2–294	0.087	0.295	<0.001
Age	0.216	<0.001					
Native tongue	4.412	<0.001					
Outcome variable: Casper	Regression B coefficient	P	F	df	R^2	R	p
Initial model			10.576	5–291	0.154	0.392	<0.001
Age	0.027	0.650					
Native tongue	7.654	<0.001					
Existing bachelor degree	1.972	0.018					
Healthcare experience	−0.642	0.270					
One parent with university degree	0.486	0.371					
Final model			25.489	2–294	0.148	0.384	<0.001
Native tongue	7.681	<0.001					
Existing bachelor degree	2.246	0.001					

Distribution of test results and test reliability

HAM-Mid I had the lowest internal consistency of all three tests with a Cronbach's alpha of 0.55 (Table 2). The item difficulty for this test was 0.38 on average (range: 0.13–0.82) with two thirds of the candidates obtaining less than half of the points assigned to this part of the test. HAM-Mid II had a better internal consistency (Cronbach's Alpha = 0.68) and a slightly lower item-total correlation (0.27) than HAM-Mid I. The item difficulty for this test was 0.77 on average with biology questions being most difficult. Casper had the highest internal consistency with a Cronbach's Alpha of 0.72.

Validity of tests

The cognitive ability tests HAM-Mid I and II correlated weakly but significantly with each other as well as with Casper performance (Table 2).

The distribution of both outcome criteria used to calculate predictive validity was quite asymmetrical with many students obtaining very good results: the science test had a median of 6.0 (skewness: −1.25; kurtosis: 0.65) while the midwifery exam had a median of 5.7 (skewness: −1.43; kurtosis: 2.12). Nevertheless, we found one significant correlation between selection criteria and students' academic performance, namely between the knowledge

Table 4
Candidates' perceptions of tests experience and appropriateness.

Theme	Quotation
General comments	"I found the test organization as well as the execution very good and uncomplicated."
	"Generally pretty smooth but intimidating. I find relying on technology very risky but that is because of the times we are living"
	"I think it's great that they have managed so well to switch to online tests."
HAM-Mid I	"The test was well structured and the questions were well-worded. However, the time frame for arithmetic problem solving is too short."
	"One minute per question is very unrealistic, you first need half a minute to read the question and make notes."
HAM-Mid II	"Unfortunately, my test stopped shortly before the end and I could no longer access it and end the process."
	"I found the selection of questions very appropriate, especially the biology and midwifery science questions."
	"Having a full lecture included was very motivating because it gave me a preview of what could be in store if I passed the test."
Casper	"Reading the subtitles was very stressful, German videos have to be made."
	" It is a great supplement to the HAM-Mid. I am very pleased that social competence is also looked at and not only the GPA mark."

admission test HAM-Mid II and the first semester oral midwifery exam (Table 2).

Relationship between socio-demographic characteristics and admission test results

The multiple regression analysis indicated that a series of socio-demographic factors such as age, previous academic achievement and native tongue were related to applicants' performance on different admission tests (Table 3). Younger candidates and those whose first language was other than German had lower HAM-Mid II scores than their peers with these two factors explaining 8.7% of the variance ($p < 0.001$). Higher Casper test scores were associated with having a bachelor degree and being a German native speaker, which explained almost 15% of the variance ($p < 0.001$). None of the considered socio-demographic factors were associated with applicants' performance in the HAM-Mid I test.

Test acceptability amongst candidates

Candidates were generally satisfied with the organisation and content of admission procedures (Table 4). Although applicants found HAM-Mid questions appropriate (Table 4), HAM-Mid I questions (on arithmetic problem solving) were considered less understandable (3.1, range: 1–5) and less fair (3.3, range: 1–5) than HAM-Mid II questions (fairness 4.1, understandable questions 4.3). Casper was deemed a good alternative to interviews and an appropriate balance to cognitive abilities criteria. It had an overall evaluation of 7.7 (SD = 1.7; range 1–10). However, reading the subtitles was also stressful for some candidates who believed that the videos should have been made with German-speaking actors. Frequently named themes mentioned by applicants in the evaluation survey were insufficient time for the HAM-Mid I speed test (mentioned 33 times) and Casper (mentioned by 44 applicants): 90.9% ($N = 221$) of candidates would have needed more time for HAM-Mid I compared to only 2.1% for the biology and midwifery questions in HAM-Mid II.

Discussion

To our knowledge, this is the first study on the validation of both cognitive ability tests and personal and professional characteristics assessment in the context of midwifery programmes selection. We developed a multiple-choice test (HAM-Mid) for cognitive abilities and administered an adapted version of the online test Casper for personal and professional characteristics.

We initially designed a paper-pencil HAM-Mid test for the selection of midwifery students which, due to the pandemic, had to be carried out online. We were not able to supervise applicants during these two tests, therefore we cannot state with certainty that the test results reflect applicants' „real abilities". However,

both tests had a time limit and therefore there were few opportunities to resort to learning material and to fake results. Moreover, HAM-Mid I proved to be very difficult in spite of theoretically having the opportunity to take notes while HAM-Mid II could still discriminate between good and poor performances. Candidates that would have otherwise not been able to travel for the tests saved time and money and were able to participate in the selection procedures. In spite of challenges posed by online selection, which have been previously acknowledged in the literature (Cleland et al., 2020), we believe that it can contribute to widening participation of a more diverse population of students to midwifery programmes – a priority in all health professions (Simone et al., 2018).

The internal consistency of the HAM-Mid I was quite low, probably a result of the small sample size and/or inadequate item difficulty. This mirrored applicants' perception of the test who considered it inappropriate, too complex and irrelevant for the midwifery profession. In order to improve its reliability as well as its acceptability, more and easier questions could be developed. Unlike a math test used to select nursing students (Underwood et al., 2013), the HAM-Mid I did not correlate significantly with study success in the first semester. However, our test was meant to measure fluid intelligence and not math skills. For future selection procedures, different alternatives to HAM-Mid I should be considered including the use of basic math or numeracy tests which are considered very relevant for clinical practice and patient safety (Roykenes and Larsen, 2010; Eastwood et al., 2011; Arkell and Rutter, 2012; Taylor et al., 2014).

The HAM-Mid II exhibited an acceptable to good internal consistency; because several different subjects were addressed, we did not expect a much higher Cronbach's alpha. The test consisted of more and easier questions than HAM-Mid I, leading to a broad distribution and sufficient variance of results. The applicants deemed HAM-Mid II better suited for midwifery selection than HAM-Mid I and considered its topics relevant and appropriate. In contrast to HAM-Mid I, the results of the HAM-Mid II were worse for non-native German speakers and younger applicants. Further analyses should clarify whether the HAM-Mid questions are systematically biased with respect to age and language. However, congruent with findings from other fields such as nursing (Newton et al., 2007; Underwood et al., 2013; Wambuguh et al., 2016) or medicine (Meyer et al., 2019) showing that knowledge tests were predictive of students' early academic performance, HAM-Mid II was the only criterion that correlated significantly with students' academic performance in a first semester midwifery exam. This correlation might be partly explained by the fact that, in contrast to the other admission criteria, HAM-Mid II was not used for the actual selection of students and therefore its variance in the subgroup of admitted students was not reduced through the ranking process. Another possible explanation for this positive correlation is that the constructs and content addressed in HAM-Mid II and the midwifery exam were similar.

Casper showed the highest internal consistency of all applied tests in spite of the fact that it addresses a very broad construct. Although intended to measure personal and professional characteristics (Dore et al., 2009), the Casper results correlated with all cognitive admission criteria. Previous studies in medical education showed a correlation of Casper results with outcome criteria that measured similar aspects of professionalism to Casper (Dore et al., 2017). However, our applicants reported in the evaluation survey on being challenged reading the German subtitles while watching the videos in English and this perception reflected in the lower Casper scores of applicants who were not native German speakers. Therefore, it is possible that the cognitive load of tasks in our sample may have been higher than for English-speaking cohorts of candidates and even higher for our non-native speaking applicants who had to mentally translate both the scenario dialogues and the subtitles into their native language and then type their responses in German.

Our aim of using Casper was to identify students with appropriate personal and professional characteristics. Due to the COVID-19 pandemic, in the first semester all exams measuring personal characteristics were cancelled so that no adequate study results for the validation of the Casper were available. Coherent with previous studies who showed the absence of a correlation between Casper and cognitive outcome criteria (Dore et al., 2017), the academic performance in the first semester did not correlate with midwifery candidates' Casper scores. Further research is needed to investigate whether tests such as Casper are able to predict midwifery students' personal and professional characteristics. Moreover, because Casper measures behavioural intention, it should be used as elsewhere only as a preselection tool and preferably complemented by multiple mini interviews. Due to their good psychometric properties (Callwood et al., 2014; Gale et al., 2016; Callwood et al., 2018) as well as their allowing the observation of actual behaviour, multiple mini interviews are preferred by midwifery stakeholders.

Our applicants' population was quite homogeneous concerning their first language, in their vast majority German native speakers. The proportion of non-native speaking students upon admission was lower than upon registration, which could indicate that our tests may unintentionally exclude certain students subgroups. Moreover, our analysis showed that candidates' first language was associated with HAM-Mid II and Casper results with native speakers obtaining better scores. However, the collected data referred to applicants' first language and not to their level of fluency in the German language, which may help us understand better some of these results. Fluency in the study language is a complex issue, which needs to be carefully pondered by educators. On the one hand, students need a certain level of language fluency that enables them to perform well academically; on the other hand, diversity of the healthcare workforce is becoming increasingly important for the quality of healthcare (Cohen et al., 2002; Grainger, 2006; Snyder et al., 2018). Educators should therefore consider strategies to facilitate access of underrepresented groups to undergraduate education (Al-Jabir, 2018) including translating test tasks into plain language containing simpler vocabulary or creating separate quotas for students whose first language is other than German. For such students additional support in the form of crash courses and language classes should be offered.

Limitations

Our study has a number of limitations. Firstly, study participants are represented by one cohort of applicants from one institution. Secondly, because some of the first semester exams were cancelled, we were unable to determine whether our admission criteria can predict academic success in terms of students' personal and professional characteristics, which are of paramount importance in midwifery education. Thirdly, our research was focused on early academic performance; future studies should investigate which cognitive abilities and personal characteristics tests contribute to retention and attrition rates of midwifery students (Green and Baird, 2009; Sabin et al., 2012; Rodgers et al., 2013). Finally, the current study presents predictive validity evidence for early academic success; however, we are currently planning a longer-term follow-up, which will allow us to observe academic performance until the end of the midwifery bachelor degree.

Conclusions

The reliability and predictive validity of the newly developed assessment methods give indications for the appropriateness of the applied selection criteria. The HAM-Mid II reliably measures knowledge in midwifery, biology and social sciences and is able to predict cognitive study success. The HAM-Mid I had a low internal consistency but was not associated with socio-demographic factors; in order to improve its reliability and validity, easier questions or math tests should be used. For future Casper administrations, an all-German content should be developed to reduce the cognitive load on applicants and to better ensure that the scenario content matches the cultural context. Finally, two of the conducted tests (HAM-Mid II and Casper) were highly accepted by applicants. Attention has to be drawn however to widening participation of underrepresented groups for a diverse future workforce.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Oana R. Groene: Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft. **Mirjana Knorr:** Conceptualization, Project administration, Methodology, Formal analysis, Writing – review & editing. **Daniela Vogel:** Conceptualization, Methodology, Writing – review & editing. **Carmen Hild:** Resources, Writing – review & editing. **Wolfgang Hampe:** Conceptualization, Methodology, Writing – review & editing.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.midw.2021.103245.

References

Al-Jabir, H.M., 2018. Widening participation: the value of diversity. Clin. Teach. 15 (4), 348–349.

Arkell, S., Rutter, P.M., 2012. Numeracy skills of undergraduate entry level nurse, midwife and pharmacy students. Nurse Educ. Pract. 12 (4), 198–203.

Bovermann, Y., 2020. [The shift to academic degree level of the midwifery profession (Part 1): opportunities - and how they can be used in the best possible way in the study programmes]. Z. Geburtshilfe Neonatol. 224 (3), 124–129.

Callwood, A., Cooke, D., Allan, H., 2014. Developing and piloting the multiple mini-interview in pre-registration student midwife selection in a UK setting. Nurse Educ. Today 34 (12), 1450–1454.

Callwood, A., Cooke, D., Bolger, S., Lemanska, A., Allan, H., 2018. The reliability and validity of multiple mini interviews (MMIs) in values based recruitment to nursing, midwifery and paramedic practice programmes: findings from an evaluation study. Int. J. Nurs. Stud. 77, 138–144.

Callwood, A., Groothuizen, J.E., Allan, H.T., 2019. The "values journey" of nursing and midwifery students selected using multiple mini interviews; year two findings. J. Adv. Nurs. 75 (5), 1074–1084.

Callwood, A., Groothuizen, J.E., Lemanska, A., Allan, H., 2020. The predictive validity of multiple mini interviews (MMIs) in nursing and midwifery programmes: year three findings from a cross-discipline cohort study. Nurse Educ. Today 88, 104320.

Carolan, M.C., Kruger, G.B., 2011. Concerns among first year midwifery students: towards addressing attrition rates. Contemp. Nurse 38 (1–2), 139–147.

Cattell, R.B., 1967. The theory of fluid and crystallized general intelligence checked at the 5-6 year-old level. Br. J. Educ. Psychol. 37 (2), 209–224.

Cleland, J., Chu, J., Lim, S., Low, J., Low-Beer, N., Kwek, T.K, 2020. COVID 19: designing and conducting an online mini-multiple interview (MMI) in a dynamic landscape. Med. Teach 42 (7), 776–780.

Cohen, J.J., Gabriel, B.A., Terrell, C., 2002. The case for diversity in the health care workforce. Health Aff. 21 (5), 90–102 (Millwood).

Dore, K.L., Reiter, H.I., Eva, K.W., Krueger, S., Scriven, E., Siu, E., Hilsden, S., Thomas, J., Norman, G.R., 2009. Extending the interview to all medical school candidates–computer-based multiple sample evaluation of noncognitive skills (CMSENS). Acad. Med. 84 (10 Suppl), S9–12.

Dore, K.L., Reiter, H.I., Kreuger, S., Norman, G.R., 2017. CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. Adv. Health Sci. Educ. Theory Pract. 22 (2), 327–336.

Eastwood, K.J., Boyle, M.J., Williams, B., Fairhall, R., 2011. Numeracy skills of nursing students. Nurse Educ. Today 31 (8), 815–818.

Effland, K.J., Hays, K., 2018. A web-based resource for promoting equity in midwifery education and training: towards meaningful diversity and inclusion. Midwifery 61, 70–73.

Farley, C., Carr, K.C., 2003. New directions in midwifery education: the master's of science in midwifery degree. J. Midwifery Womens Health 48 (2), 133–137. doi:10.1016/s1526-9523(02)00423-3.

Ferguson, E., James, D., Madeley, L., 2002. Factors associated with success in medical school: systematic review of the literature. BMJ 324 (7343), 952–957.

Gale, J., Ooms, A., Grant, R., Paget, K., Marks-Maran, D., 2016. Student nurse selection and predictability of academic success: the multiple mini interview project. Nurse Educ. Today 40, 123–127.

Grainger, K., 2006. Equal access to training for black and minority ethnic nurses. Nurs. Stand. 20 (42), 41–49.

Green, S., Baird, K., 2009. An exploratory, comparative study investigating attrition and retention of student midwives. Midwifery 25 (1), 79–87.

Hakimi, M., Kheirkhah, M., Abolghasemi, J., Hakimi, R., 2021. Investigating the effect of neonatal resuscitation simulation using a competency-based approach on knowledge, skill, and self-confidence of midwifery students using objective structured clinical examination (OSCE). J. Fam. Med. Prim. Care 10 (4), 1766–1772.

Hermansson, E., Mårtensson, L.B., 2013. The evolution of midwifery education at the master's level: a study of Swedish midwifery education programmes after the implementation of the Bologna process. Nurse Educ. Today 33 (8), 866–872. doi:10.1016/j.nedt.2012.09.015.

Hissbach, J.C., Sehner, S., Harendza, S., Hampe, W., 2014. Cutting costs of multiple mini-interviews - changes in reliability and efficiency of the Hamburg medical school admission test between two applications. BMC Med. Educ. 14, 54.

Hughes, H.A., 2013. Factors influencing attrition rates in midwifery students. Nurs. Stand. 27 (26), 42–48.

Malakooti, N., Bahadoran, P., Ehsanpoor, S., 2018. Assessment of the midwifery students' clinical competency before internship program in the field based on the objective structured clinical examination. Iran J. Nurs. Midwifery Res. 23 (1), 31–35.

Meyer, H., Zimmermann, S., Hissbach, J., Klusmann, D., Hampe, W., 2019. Selection and academic success of medical students in Hamburg, Germany. BMC Med. Educ. 19 (1), 23.

Newton, S.E., Smith, L.H., Moore, G., Magnan, M., 2007. Predicting early academic achievement in a baccalaureate nursing program. J. Prof. Nurs. 23 (3), 144–149.

Patterson, F., Zibarras, L., Ashworth, V., 2016a. Situational judgement tests in medical education and training: research, theory and practice: AMEE guide no. 100. Med. Teach. 38 (1), 3–17.

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., Cleland, J., 2016b. How effective are selection methods in medical education? A systematic review. Med. Educ. 50 (1), 36–60. doi:10.1111/medu.12817.

Pitt, V., Powis, D., Levett-Jones, T., Hunter, S., 2014. Nursing students' personal qualities: a descriptive study. Nurse Educ. Today 34 (9), 1196–1200.

Plappert, C., Simoes, J.G., Schönhardt, S., Abele, H., 2019. The academization of midwifery in the context of the amendment of the German midwifery law: current developments and challenges. Geburtshilfe Frauenheilkd 79 (8), 854–862.

Puddey, I.B., Mercer, A., 2013. Socio-economic predictors of performance in the undergraduate medicine and health sciences admission test (UMAT). BMC Med. Educ. 13, 155.

Rodgers, S., Stenhouse, R., McCreaddie, M., Small, P., 2013. Recruitment, selection and retention of nursing and midwifery students in Scottish Universities. Nurse Educ. Today 33 (11), 1301–1310.

Roykenes, K., Larsen, T., 2010. The relationship between nursing students' mathematics ability and their performance in a drug calculation test. Nurse Educ. Today 30 (7), 697–701.

Sabin, M., Taylor, R., Tilley, C., 2012. Untangling a complex issue: an overview of initiatives to support nursing and midwifery student recruitment, selection and retention in Scottish Universities. Nurse Educ. Today 32 (4), 469–474.

Simone, K., Ahmed, R.A., Konkin, J., Campbell, S., Hartling, L., Oswald, A.E., 2018. What are the features of targeted or system-wide initiatives that affect diversity in health professions trainees? A BEME systematic review: BEME Guide no. 50. Med. Teach. 40 (8), 762–780.

Siu, E., Reiter, H.I., 2009. Overview: what's worked and what hasn't as a guide towards predictive admissions tool development. Adv. Health Sci. Educ. Theory Pract. 14 (5), 759–775.

Smith, V., Muldoon, K., Biesty, L., 2012. The objective structured clinical examination (OSCE) as a strategy for assessing clinical competence in midwifery education in Ireland: a critical review. Nurse Educ. Pract. 12 (5), 242–247.

Snyder, C.R., Frogner, B.K., Skillman, S.M., 2018. Facilitating racial and ethnic diversity in the health workforce. J. Allied Health 47 (1), 58–65.

Sweet, L., Glover, P.A., McPhee, T.A., 2013. The midwifery miniCEX-A valuable clinical assessment tool for midwifery education. Nurse Educ. Pract. 13 (2), 147–153.

Taylor, R., Macduff, C., Stephen, A., 2014. A national study of selection processes for student nurses and midwives. Nurse Educ. Today 34 (8), 1155–1160.

Underwood, L.M., Williams, L.L., Lee, M.B., Brunnert, K.A., 2013. Predicting baccalaureate nursing students' first-semester outcomes: HESI admission assessment. J. Prof. Nurs. 29 (2, Supplement 1), S38–S42.

Ventura, C.A., Mendes, I.A., Fumincelli, L., Trevizan, M.A., 2015. The evolution of world health organization's initiatives for the strengthening of nursing and midwifery. J. Nurs. Scholarsh. 47 (5), 435–445.

Wambuguh, O., Eckfield, M., Van Hofwegen, L., 2016. Examining the importance of admissions criteria in predicting nursing program success. Int. J. Nurs. Educ. Scholarsh. 13 (1), 87–96.

Waugh, A., Smith, D., Horsburgh, D., Gray, M., 2014. Towards a values-based person specification for recruitment of compassionate nursing and midwifery candidates: a study of registered and student nurses' and midwives' perceptions of prerequisite attributes and key skills. Nurse Educ. Today 34 (9), 1190–1195.

Webster, E.S., Paton, L.W., Crampton, P.E.S., Tiffin, P.A., 2020. Situational judgement test validity for selection: a systematic review and meta-analysis. Med. Educ. 54 (10), 888–902.

World Health Organization, 2016. Standards For Improving Quality of Maternal and Newborn Care in Healthcare Facilities. World Health Organization, Geneva.

World Health Organization, 2017. Nursing and Midwifery in the History of the World Health Organization 1948-2017. World Health Organization, Geneva.

Yigzaw, T., Ayalew, F., Kim, Y.M., Gelagay, M., Dejene, D., Gibson, H., Teshome, A., Broerse, J., Stekelenburg, J., 2015. How well does pre-service education prepare midwives for practice: competence assessment of midwifery students at the point of graduation in Ethiopia. BMC Med. Educ. 15, 130.

Zamanzadeh, V., Ghahramanian, A., Valizadeh, L., Bagheriyeh, F., Lynagh, M., 2020. A scoping review of admission criteria and selection methods in nursing education. BMC Nurs. 19 (1), 121.

International Confederation of Midwives, 2ß19. Essential competencies for midwifery practice - 2019 update. Retrieved from: https://www.internationalmidwives.org/our-work/policy-and-practice/essential-competencies-for-midwifery-practice.html.