Report to the Nursing and Midwifery Council on Language Testing Policy

Anthony Green, Sathena Chan
University of Bedfordshire

June 2022

Executive summary

Responding to the NMC's review of its language testing policy, our project involved:

- A review of the extent to which the approach to language testing currently adopted by the NMC is proportionate and appropriate, and
- Recommendations for a methodology to investigate whether language tests of interest should be accepted by the NMC

Our report:

- Focusses on the recognition of language tests but also considers the basis for registration via periods of training, clinical practice, or through appeal
- Indicates that the NMC's current language testing policy is broadly in line with international practice, but should be given firmer theoretical foundations
- Recommends methodology changes, expanding the NMC's current criteria for recognising evidence for the English language abilities of applicants for registration

Recommendations:

- NMC policy should be evaluated against established professional standards for language assessment
- NMC policy should take greater account of the gap between what language tests are able to assess and the practical use of language in clinical and professional development settings, indicating the need for…
    - ongoing support for internationally educated professionals working in the UK
    - consideration of a wider range of options for evidencing English language abilities
- Recognition by the NMC of evidence for English language abilities should be based on the framework we propose
- Assessment providers or sponsors should present their case for recognition based on requirements set out by the NMC
- Continuing recognition should be made contingent on periodic reviews to ensure that NMC requirements continue to be met

Table of Contents

1. Overview

Internationally trained professionals on the NMC register make a central contribution to nursing and midwifery in health and social care settings across the UK. Approximately 17.5% of those currently listed on the register (131,640 individuals) received their nursing education outside the UK. Data released in May 2022 shows that in the period April 2021 to March 2022, 23,408 internationally trained nurses, midwives and nursing associates joined the register compared with 25,028 UK trained.

Effective communication is essential to healthcare work. As many internationally educated nurses and midwives (IENMs) are speakers of languages other than English, the NMC requires applicants for registration to evidence a sufficient level of English language ability to support safe and effective practice. This report offers an evaluation of the current NMC language assessment policy. We were also invited to consider whether alternative language tests should be accepted by the NMC (in addition to the two currently recognised) but have chosen to recommend criteria for the recognition of language assessments and a process of evaluation that can be applied to any form of assessment that might be proposed.

1.1. Methods

To support our review of the extent to which the approach to language assessment currently adopted by the NMC is proportionate and appropriate, we were provided with a number of NMC documents including both published materials and internal documents. These documents included policy guidance, briefing papers, records of consultations and training materials connected to the NMC's current language assessment policy. Throughout the project, we attended regular meetings with the NMC where we were able to raise questions, seek clarification and request additional documentation.

We also reviewed the NMC website (www.nmc.org.uk) and, for purposes of comparison, the websites of other nursing and midwifery regulators in English speaking countries including Australia (www.nursingmidwiferyboard.gov.au, www.ahpra.gov.au), Canada (www.nnas.ca, websites for nursing colleges in individual provinces), Ireland (www.nmbi.ie), New Zealand (www.nursingcouncil.org.nz, www.midwife.org.nz) and the USA, (www.ncsbn.org, websites for individual state boards of nursing).

We consulted the recent language assessment literature to identify sets of standards for assessment that are widely cited in language test validation studies. We used these standards and recent projects (notably Chapelle, Enright and Jamieson 2008) as the basis for our review of current NMC practices and our proposals for evaluation criteria.

We then carried out a rapid review of literature relating to the language demands and registration requirements for nurses, midwives, and nursing associates in English speaking countries. The purpose of this review was to collect evidence and opinions that might support or challenge current NMC practices. To identify studies for inclusion, we conducted a search of the ProQuest *British Nursing Database* and EBSCO *CINAHL Complete* and *Education Research Complete* databases using the following search terms: *nurs**, *midw**, *English language*. We limited the search to the years 2002 to 2022 inclusive and to papers published in English. In this way, we identified 182 potentially relevant papers and although we were unable to carry out a

systematic thematic analysis within the scope of the project, we read through the abstracts or introductory sections of these to identify for fuller reading those that appeared most relevant to the questions we were asked to address. Through the websites of language assessment organisations whose tests were recognised by one or more of the regulators (www.ielts.org, www.occupationalenglishtest.org, www.celbancentre.ca, www.ets.org, michiganassessment.org, www.pearsonpte.com), we were able to locate an additional ten grey literature sources relating to the use of tests for nursing or midwifery registration.

Finally, we prepared a set of criteria and recommendations for a regular evaluative process that would give the NMC a consistent and defensible basis for the recognition of language assessments for registration purposes. We then held a session with NMC staff to obtain feedback on these proposals and interviewed two experts in healthcare-related language education and assessment, incorporating their suggestions into this report.

1.2. Current NMC policy

It is a prerequisite for registration with the NMC that all applicants should be able to be "communicate clearly and effectively in English" (NMC 2018b). Under current regulations, this ability may be evidenced in one of three ways:

- Recent achievement of the required score in one of the English language tests accepted by the NMC.
    - The NMC currently lists the Academic variant of the International English Language Test System (IELTS) and the Occupational English Test (OET) as accepted English language tests. To be considered for registration, applicants require a score of at least 7.0 on the Speaking, Listening and Reading and 6.5 on the Writing component of the IELTS, or a at least a grade B in the reading, listening and speaking sections and a grade of at least C+ in the writing section of the OET.
    - In its Guidance on registration language requirements (NMC 2020) the NMC sets seven requirements for accepting English language tests:
      22.1 it tests knowledge of English in either a healthcare or academic context. It must not be a general test;
      22.2 it tests reading, writing, listening and speaking;
      22.3 the speaking element is tested face-to-face and not via a computer test;
      22.4 there are high level security procedures for the test production and delivery;
      22.5 there are test score verification systems that allow the NMC to confirm your results;
      22.6 there is appropriate evidence of the reliability of the test and the dependability of its scores; and
      22.7 there are test centres widely available.
- A pre-registration nurse, midwife or nursing associate qualification taught and examined in English.
- Recent practice of at least one year in a majority English speaking country.

1.3. Appeals

Additionally, decisions based on the three forms of evidence listed above are open to appeal. Applicants who fail to meet the requirements but feel that they do have the appropriate English language standard to practice safely and effectively may ask to have their case heard by an independent panel (NMC 2021a, 2021b, n.d.).

We considered the appeals process to be a weakness of the current approach. Our chief concern is that the process invites appeals panellists, who may have no background in language education, to use their own judgement to make decisions about the adequacy of the appellant's language skills based on their performance at the hearing, and to judge whether a test other than IELTS and OET might be accepted as alternatives (NMC n.d.). We observe that an appeals hearing does not include procedures explicitly designed to a) elicit and assess language abilities or b) to evidence the qualities of language assessments, and that panellists may not be best placed to judge these. The appeals process would be strengthened if it could be demonstrated that it follows the same underlying principles that inform acceptance of the standard forms of evidence.

2. Standards for language testing

Regarding the ability of IENMs to communicate in English, the Nursing and Midwifery Council needs to find a balance between two risks, each of which could jeopardise safe and effective practice. One is the risk that communication with colleagues and patients may be inadequate; the other, the risk that requirements for high levels of English language proficiency may result in the registration of too few nurses, midwives or nursing associates to support safe and effective care.

In considering whether the NMC language assessment policy is proportionate and appropriate, the researchers were informed by industry standards and guidelines produced by specialists in assessment. These included the *Standards for Educational and Psychological Testing* published by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (AERA/APA/NCME 2014). Although developed for use in North America, the *AERA/APA/NCME Standards* are often consulted internationally and offer comprehensive "criteria for the development and evaluation of tests and testing practices and… provide guidelines for assessing the validity of interpretations of test scores for the intended test uses" (p.1). We also referred to the International Language Testing Association's (ILTA) (2007) *Guidelines for Practice*, and the Association of Language Testers in Europe (ALTE) (1994) *Code of Practice.* Both are compatible with the *AERA/APA/NCME Standards* and considerably less comprehensive. However, these two documents, unlike the *AERA/APA/NCME Standards,* are specifically concerned with language assessments, the focus of this paper. Extracts from the three documents relating to assessment interpretation and use are reproduced in Box 1.

Among the *AERA/APA/NCME Standards*, Standard 9 is of particular relevance to the NMC as it concerns the rights and responsibilities of test users: "those who use the test results for some decision-making purpose (including policy makers and those who use data to inform social policy)". (p.3). The first sentence of the overarching standard (Standard 9.0) states,

"Test users are responsible for knowing the validity evidence in support of the intended interpretations of scores on tests that they use, from test selection through the use of scores, as well as common positive and negative consequences of test use." (p.142).

Similarly, the *ILTA Guidelines* list five "Responsibilities of users of test results":

"Persons who utilize test results for decision making must:

1. *Use results from a test that is sufficiently reliable and valid to allow fair decisions to be made.*
2. *Make certain that the test construct is relevant to the decision to be made.*
3. *Clearly understand the limitations of the test results on which they will base their decision.*
4. *Take into consideration the standard error of measurement (SEM) of the device that provides the data for their decision.*
5. *Be prepared to explain and provide evidence of the fairness and accuracy of their decision making process."* (p.5)

The *ALTE Code of Practice* lists four "responsibilities of examination users". These include,

*"**Selecting Appropriate Examinations***

Examination users should select examinations that meet the purpose for which they are to be used and that are appropriate for the intended candidate populations.

Interpreting Examination Results

Examination users should interpret scores correctly.

Striving for Fairness

Examination users should select examinations that have been developed in ways that attempt to make them as fair as possible for candidates of different backgrounds (e.g. gender, ethnic origin, special needs, etc.).

Informing Examination Takers

In cases where the examination user has direct communication with candidates, they should regard themselves as having many of the obligations set out for [test providers] under the section in Part One entitled Informing Examination Takers." (p.5).

The three sets of guidelines quoted in Box 1 all suggest that the NMC, as a user of assessment results (which we take to include all three accepted forms of evidence of English language abilities, as well as the appeals process) has a responsibility, shared with the organisations whose assessments it recognises, to justify its acceptance of these forms of evidence and its

rejection of alternatives. This includes demonstrating to interested parties how each form of evidence supports fair and equitable decision making.

2.1. Validation arguments

The guidelines we consulted (Box 1) point to the importance of three related elements:

1. the forms of evidence accepted (test scores, training received in English, practice in an English-speaking country),
2. the interpretation made of that evidence (that an applicant has sufficient language ability to support safe and effective practice in the UK) and
3. the likely consequences of using the selected forms of evidence. Consequences may be positive (such as enhanced levels of patient safety, encouraging applicants to improve their language skills); or negative (such as discouraging qualified individuals from applying; imposing unwarranted costs on applicants; disadvantaging members of certain social groups).

To justify the use of an assessment in the face of potential legal challenges, current practice involves the development of 'validity arguments'. Grounded in the work of Toulmin (1969), these involve articulating a series of linked claims that connect evidence provided by an assessment to the interpretation and (intended or anticipated) consequences of using that assessment. Kane (2013), a leading proponent of argument-based validation, explains,

> *We generally do not employ test scores simply to report how a test taker performed on certain tasks on a certain occasion and under certain conditions. Rather, the scores are used to support claims that a test taker has, for example… some probability of succeeding in an educational program or other activity. These claims are not generally self-evident and merit evaluation.*
>
> *To validate an interpretation or use of test scores is to evaluate the plausibility of the claims based on the test scores. Validation therefore requires a clear statement of the claims inherent in the proposed interpretations and uses of the test scores. Public claims require public justification.* (Kane 2013, p.2).

The researchers set out to outline and evaluate the claims that the NMC might wish to make regarding the interpretation of the forms of assessment it accepts for nursing and midwifery registration. We first reviewed documents provided by the NMC to identify claims that are made or implied in articulating the English language policy.

To give an example, one key claim (relevant to point 2 of the *ILTA Guidelines* and the first point listed from the *ALTE Code of Practice*) would be that the language abilities required for safe and effective practice as a nurse, midwife or nursing associate are adequately represented and measured by three accepted forms of assessment (i.e., scores on the approved language tests, completion of English-medium training, or a period of practice in an English-speaking country).

In constructing a validity argument, it is not only necessary to explain the assumptions behind and grounds for accepting each form of evidence, but also to state and evaluate rebuttals: potential counterclaims that might call the validity of the NMCs interpretation of the evidence into question. An example of such a rebuttal aimed at the current NMC policy would be the allegations that have sometimes been made (for example by Pilcher and Richards 2017) that

success on the English language tests accepted by the NMC requires language abilities that are of limited relevance to the practice of nursing.

The researchers carried out a rapid review of relevant literature to identify both support for, and challenges to the plausibility of the NMC's claims. In the following sections, we use the steps in the validation argument presented for the iBT TOEFL by Chapelle, Enright and Jamieson (2008) of Domain Description, Evaluation, Generalization, Explanation, Extrapolation, and Utilization. We evaluate the evidence from the NMC and the literature supporting or challenging the validity of inferences based on each of the three forms of assessment currently accepted by the NMC. We conclude with outline proposals for a practical framework for the evaluation of assessments of English language abilities for purposes of professional registration.

3. Domain Description

This step in a validation argument involves the connection between the language skills required in the target domain (of nursing/ midwifery practice) and the language tested by the assessment. The claim that the NMC policy implies might be worded as follows.

Claim:

Performance of tasks on the assessment adequately reflects the language knowledge, skills, and abilities required for safe and effective practice in the target domain of language use: nursing/ midwifery practice in the UK.

3.1. English language abilities in the NMC Code

The NMC *Code* (NMC 2018b) provides guidance on the language abilities required for safe and effective communication in English on the part of IENMs. This is most explicit in Section 7, which presents five statements explaining what it means for a nurse, midwife or nursing associate to communicate clearly:

7.1 use terms that people in your care, colleagues and the public can understand
7.2 take reasonable steps to meet people's language and communication needs, providing, wherever possible, assistance to those who need help to communicate their own or other people's needs
7.3 use a range of verbal and non-verbal communication methods, and consider cultural sensitivities, to better understand and respond to people's personal and health needs
7.4 check people's understanding from time to time to keep misunderstanding or mistakes to a minimum
7.5 be able to communicate clearly and effectively in English

English language abilities are also implicated in several other statements in the Code (NMC 2018b). For example, in Sections 9 "Share your skills, knowledge and experience for the benefit of people receiving care and your colleagues", 10 "Keep clear and accurate records relevant to your practice", and in statements such as 22.3 "keep your knowledge and skills up to date, taking part in appropriate and regular learning and professional development activities that aim to maintain and develop your competence and improve your performance."

Taken together, these statements provide an outline domain description that could be used to evaluate the appropriateness of language assessments for the purpose of registration. Studies such as Sedgwick and Garner (2017) provide language data that adds detail on how nurses and midwives use language in their work and professional study.

3.2. Evidence types and domain description

Inferences based on the second and third of the NMC evidence types are relatively direct. Applicants who have completed training through English (including a period of clinical practice) or practised in an English-speaking environment have been given scope to demonstrate their ability to communicate in the target domain over an extended period.

However, acceptance of these two forms of evidence would seem to rest on a number of assumptions that could be questioned. One is that practice in other English-speaking healthcare settings is broadly comparable to the target domain of practice in the UK. While there are differences internationally in nursing and midwifery practice (and in wider cultural practices), there is little evidence to suggest that these have a substantial effect on safety or effectiveness (although this question does not appear to have been extensively researched). The policy also assumes that the mandated periods of study or practice are adequate to establish whether a practitioner may have difficulty in understanding and communicating through English. Additionally, the policy assumes that any such difficulties would be consistently detected and acted on. Applicants who lacked adequate English language abilities and failed to meet the requirements of the Code (NMC 2018b) would fail their courses or be denied the continuing right to practice. These latter questions are taken up in Section 4.

Inferences based on language test scores are more open to challenge on domain definition grounds because test performance, even if based on the communication needs suggested by the Code (NMC 2018b), cannot directly reflect language use in the target domain of professional practice. The current NMC policy (NMC 2020, para.22.1) embraces a "languages for specific purposes" (Douglas 2000) view of language as partly context dependent. The policy recognises tests that assess English in a 'healthcare' or 'academic' context, but not those that do so within a 'general' context. It also requires (para.22.2) that the traditional 'four skills' of reading, writing, listening and speaking should all be tested, presumably because all are routinely needed to fulfil the various purposes listed in the Code (NMC 2018b). We note that the requirement is simply to *test* the four skills, but not to report scores for each of them. No guidance is given on the minimum acceptable extent of testing of each skill.

The NMC (2020) guidelines on test recognition (22.3) require that "the speaking element is tested face-to-face and not via a computer test". Given that current NMC policy includes recognition of the '@home' and 'on computer' versions of OET and the 'computer based IELTS Academic' (all of which involve the use of a computer by test takers), the wording of this guideline could be misleading. It may be that the NMC believes that language abilities needed for spontaneous interaction are not adequately assessed through tasks that involve responding to pre-recorded questions delivered by a machine. The requirement and the rationale for it should be clarified, particularly as it excludes (and so might be challenged by) providers of a number of tests that are recognised in other jurisdictions. At the same time, because the '@home' option allows for the administration of OET outside traditional test centres NMC

(2020) Guideline 22.7 ("there are test centres widely available") appears to be losing its relevance. It could prove to be an unnecessary barrier to the recognition of innovative test providers seeking to dispense with increasingly outmoded networks of physical test centres.

While the reasons for accepting tests set in healthcare contexts may appear self-evident, no definition or description of the language requirements for nurses, midwives or nursing associates is presented or referred to in the NMC guidance and there is no specification of the kinds of task that might make a test suitable for its purpose: what it means for a test to cover "the necessary knowledge of English" in the words of Article 33 of the Health Care and Associated Professions (Knowledge of English) Order, 2015. NMC guidance is lacking on the kinds of reading, writing, listening and speaking activities that might represent suitably 'academic' or 'healthcare' contexts rather than unsuitably 'general' contexts. This delegates responsibility for domain definition and the case for the relevance of test content to the test providers. The same is true for the academic option, although this has attracted more criticism because academic language use is not directly or obviously linked to practice.

Of the tests currently recognised by the NMC, the one labelled as 'academic' – IELTS Academic – includes the same Listening and Speaking components as the IELTS General Training test, which is not accepted. Yet both tests were designed in part to address educational contexts and both tests balance more academic with more socially oriented content. The shared Listening test involves social conversations as well as short lectures; the shared Speaking test involves exchanging personal information and discussion.

The OET is promoted as a test for healthcare professionals, being based on analyses of language use in healthcare settings (see for example Knoch et al. 2020, Seguis & Lim 2020) and with a variant that explicitly targets nursing (in two of its four components). The case for the relevance of the IELTS test construct to the purpose of professional registration is less evident. It is not clearly set out on the IELTS website. The site does carry a claim that the test "is suitable for those wanting to study in an English-speaking environment or university (higher education)" but simply adds that "You can also take IELTS Academic for professional registration purposes." (www.ielts.org/about-ielts/ielts-test-types). This avoids stating that IELTS is "suitable" for such purposes and does not suggest a strong case for the test's relevance to healthcare. A number of critics, including many IENMs, have questioned the appropriateness of the current version of the test for registration (see for example Macqueen et al. 2016, Pilcher & Richards 2017, Müller & Daller 2019, Baldwin & Cheng 2020, Bond et al. 2020, Carr 2021). It appears that the convenience and ready availability of IELTS encouraged its acceptance. Its use "across Government in other domains (such as by Home Office for visa applications)" was cited as a reason for its continuing recognition in the meeting of Council on 8 October 2015, but this widespread use could equally be taken as an indication of the test's 'general' nature, arguing against its recognition. As more obviously healthcare-oriented alternatives such as OET and CELBAN have become available, the case for the continuing reliance on IELTS has become less persuasive.

Several researchers have argued that IELTS is less suitable for use in healthcare settings than OET, or other tests (Müller & Daller 2019, Baldwin & Cheng 2020, Carr 2021). One study conducted for the British Council by Sedgwick, Garner and Vicente-Macia (2016, Sedgwick and

Garner 2017) compared the language needs of practising nurses in the UK (identified by tracking four nurses and conducting two focus groups) with the content of IELTS, concluding that there were, “a number of areas where the IELTS test does not seem to assess adequately and appropriately nurses’ linguistic preparedness for practice, particularly with regard to pragmatic and social aspects of language.” (2016 p.35). The authors argued in favour of the development of a healthcare, or nursing-specific test for the UK, similar to the OET (not recognised by the NMC at that time), or the Canadian English Language Benchmarks Assessment for Nurses (CELBAN). This view was also reflected in responses to the NMC’s (2015) consultation on English language requirements.

However, aspects of OET and of CELBAN (although these were designed specifically for healthcare settings) have also been challenged in the literature on grounds of relevance. Mismatches have been found between test content and language use in practice. Lynch (2016), for example, included OET as well as IELTS in observing that, “Many nurse registering authorities have demonstrated concerns regarding English language testing with the academic nature of these tests being identified as a barrier to registration for these nurses in Australia (O'Connor 2008; Walters 2008; Deegan and Simkin 2010).” (p.539). Baldwin and Cheng (2020) found that the test candidates in their focus groups appreciated the relevance of the CELBAN to practice but also pointed out discrepancies, for example between the spoken role play in the test and the kinds of real-life interaction it is intended to simulate.

Ultimately, although some tests may better reflect the use of language for nursing and midwifery than others, it may not be realistic to expect any language test delivered in a standard format and on a large scale to capture the full scope and complexity of language use in healthcare settings. Reflecting the kinds of criticism made by Sedgwick and her colleagues (2016, 2017), a discussion paper by Müller (2016) compared the design of the OET and IELTS, noting “stark” differences in vocabulary between the two. Whereas OET involves, “vocabulary of common diseases, taking patient histories, discussing health-related issues, interpreting health-related research, and so on.” (p.135), IELTS covers a much wider variety of topics, few obviously connected to health. She noted that while the OET writing and speaking components are modelled on professional communication, IELTS speaking involves, “personalised and opinionated communication [that] may not be suitable in a professional setting” (p.135). The writing tasks consist of a description of a graphic or table and a short essay expressing a personal point of view.

Although considering that OET may offer more opportunities than IELTS for test takers to “demonstrate their linguistic readiness to enter the clinical setting” (p.136), unlike Sedgwick et al. (2016, 2017), Müller accepted IELTS as a relevant and useful test for nurses, noting that it had much in common with OET and (taking the more generalist view among some applied linguists that language abilities can transfer quite readily across contexts) that the core language skills assessed by both tests underpinned professional as well as academic activities.

Müller (2016) recognised that language tests, however closely modelled on a target domain, can provide only a limited picture of a candidate’s ability to communicate in real-life settings. She argued that "What is established by these tests… is that the candidate has the necessary foundational skills to produce and comprehend English at a sufficient level to begin the long

and continuing journey of developing their clinical communication skills. Having a language skills standard ensures that potential registered nurses commence with the same minimum core language skills." (p.136). The weakness of Müller's position as support for the current NMC policy of differentiating between healthcare, academic and general tests is that it leaves the necessary 'core language skills' under defined. It could be argued that all language tests target the "grammar, syntax, vocabulary, and fluent spontaneous speech" (p.134) listed by Müller as core competencies underlying language use in clinical settings. However, IELTS was designed with a specific purpose in mind so its claims to validity must rest at least in part on the relevance of its academic content to the specific context of healthcare.

A further argument favouring the use of healthcare-oriented testing (although associated with aspects of utilization rather than domain definition) is that it will encourage learners preparing for the test towards developing their knowledge of the language of healthcare and their skill in putting language to use in clinical contexts. There is a risk that preparing for IELTS might persuade language learners to focus instead on (for writing and speaking) expressing opinions on a wide variety of general topics (Lynch 2016).

Wette (2012) agreed with both Sedgwick et al. (2016) and Müller (2016) that IELTS and OET "do not directly assess the kind of pragmatic and discourse competence that health professionals need to master in order to communicate effectively with patients or clients." (p.109). However, her recommendation did not involve the development of new tests. Rather, she advocated that language tests should not be used as simple gatekeeping devices, but should play a more supportive role for registrants, informing the provision of targeted instruction directed at developing the "linguistic, pragmatic, discourse, and intercultural competence of overseas qualified health professionals as well as their clinical communication skills" (p.111) that are not well captured by tests. Wette suggested that lower levels of English language ability could be accepted to the extent that qualified IENMs were then offered ongoing support to promote safe and effective practice.

A more comprehensive and equitable policy might balance expectations for IENMs with expectations for ongoing help from employers for IENMs and a focus on developing effective intercultural communication skills across the healthcare sector. We note that the need for intercultural communication emerges as an important theme in the literature, including for first language practitioners working with an increasingly diverse population of patients and colleagues. In this respect, IENMs may have as much to contribute as to learn.

3.3. Domain description as a basis for test recognition

In relation to domain description, the challenge for the NMC as users of assessment results is to establish the core communication skills that IENMs need to have at registration and that suitable language tests therefore need to evidence. For any test to be accepted by the NMC, the assessment provider should be able to supply a domain description and set out a convincing case for how the tasks on the assessment address essential communication skills under operative conditions. It is also important to understand the additional clinical communication

skills that may not be directly evidenced through test performance but that often prove challenging for IENMs, including those who have successfully met the registration language requirements (Okougha & Tilki, 2010 Crawford & Candlin 2012, Stubbs 2017, Chok et al. 2018, Lin et al. 2018, Ghazal et al. 2020, McKitterick et al. 2022).

Echoing earlier studies, a survey of 110 international student nurses by McKitterick et al. (2022) found that participants experienced "difficulty comprehending Australian colloquialism, difficulty understanding different accents and lack of experience interacting with people who speak only English." (p.7). To the extent that such skills are important for safe and effective practice, they will need to be developed through further training, or addressed through other forms of supportive provision. It should be clear to employers that passing a language test (or studying or practicing for a period in an English-speaking environment) does not imply that IENMs will encounter very few challenges involving language and intercultural communication. Ethically, the responsibility for overcoming such challenges must be shared between all involved.

In a framework for evaluating language assessments (including periods of study or practice as well as tests) the rationale for linking the assessment to the domain of professional practice should be transparent.

3.4. Domain description: conclusions

Our **key conclusions** from our review of the issue of domain description can be summarised as follows:

- Some tests more closely reflect the work of nurses, midwives and nursing associates than others. For recognition, test providers should be able to articulate clearly and in detail why their test is suitable for registration purposes.
- It is unrealistic for any test to fully reflect the use of English in healthcare contexts. In the brief period available, tests need to cover common ground and cannot reasonably be expected to capture local regional, or highly specialised uses of language.
- As a result, those who meet the English language requirements (including those who do so by virtue of being first language English speakers) may nonetheless lack essential communication skills required for effective communication in healthcare environments. Practising IENMs are likely to experience challenges with issues such as locally prevalent accents, idioms and patterns of speech, cultural beliefs and values, mediating between professional and lay registers.
- Policy should acknowledge the gap between the baseline evidence that can be provided by tests and the ongoing role for English language and mutual intercultural support in promoting safety and effectiveness.

4. Evaluation and Generalization

Evaluation
Building on the domain definition, this step in the validity argument concerns how an assessment is administered and how performance is judged.

Claim:

Performance on the assessment will generate results that represent the targeted language abilities.

This assumes that

1. The conditions for administration of the assessment are appropriate for providing evidence of the English language abilities targeted by the assessment.
2. The procedures for judging or scoring performance provide suitable evidence of the English language abilities targeted by the assessment.
3. The statistical characteristics of items, measures, and test forms support consistent decision making.

Generalization
This step in the argument concerns the extent to which the evidence of English language abilities obtained through one form of an assessment is equivalent to the evidence obtained through another. Do two forms of a test involve the same abilities and are they equally difficult? Or, do two examiners judge the same performance in the same way?

Claim:

If the people being assessed were to undertake parallel versions of tasks or test forms or to be judged by different people, their results would be the same.

For language tests, there are well-established standards for test administration. The *AERA/APA/NCME Standards* cover such issues under Standard 6.0 "To support useful interpretations of score results, assessment instruments should have established procedures for test administration, scoring, reporting, and interpretation. Those responsible for administering, scoring, reporting, and interpreting should have sufficient training and supports to help them follow the established procedures. Adherence to the established procedures should be monitored, and any material errors should be documented and, if possible, corrected." (p.114). Metrics used to evaluate the consistency with which judgements are made are included under Standard 2.0: "Appropriate evidence of reliability/precision should be provided for the interpretation for each intended score use." (p.42).

Three of the NMC's current requirements for tests are closely related linked to these standards:

22.4 there are high level security procedures for the test production and delivery;
22.5 there are test score verification systems that allow the NMC to confirm your results;
22.6 there is appropriate evidence of the reliability of the test and the dependability of its scores

It is fundamental that results submitted as evidence of English language abilities by an applicant for registration should reflect that individual's own performance. Any form of malpractice such

as the use of substitute test takers, copying, leaks of test material, or misleading reports by evaluators would undermine the valid use of those results.

Any assessment provider wishing to have their assessment recognised by the NMC should therefore be able to demonstrate the quality of their security procedures and standards for assessment administration, providing the NMC with details of measures they take to monitor and assure these. Test score verification procedures can be understood as part of this requirement.

Although not covered by the *AERA/APA/NCME Standards*, assessment providers should also be able to demonstrate that, in the terms used by Ofqual (2020) their organisation has adequate "governance, integrity, resources and expertise." (p.4) to deliver quality assessments. This might be evidenced through arrangements for the regulation of the assessment provider in its home jurisdiction. Regarding reliability, in line with the *AERA/APA/NCME Standards*, test providers should estimate a conditional standard error for each skill tested at the score level used in determining whether the applicant has satisfied the English language requirement.

To support claims for generalization, test providers should be able to show that test forms derive from a common set of specifications and that they are of equivalent difficulty (also covered by Standard 2.0 in the *AERA/APA/NCME Standards*).

As noted above, an assumption underlying the acceptance of periods of study or of practice in English seems to be that these are likely to reveal any inadequacies in English language abilities and that such inadequacies are likely to result in a failure to graduate, or withdrawal of registration. In contrast to the use of language tests, the extent to which these assumptions are met – and the consistency with which English language abilities are judged through these methods – seems to have attracted little focussed attention and must be open to question. Although direct observation of performance in clinical settings satisfies requirements for domain definition, it is much less clear that the judgements applied to performance in these settings are consistent and unbiased. There have been cases of applicants falsifying documents to support applications for registration (see for example, Hackett 2021) and it is important that all methods of evidencing English language abilities should involve adequate forms of verification and monitoring.

Regulators often consider international practices when setting policy. Although sometimes criticised for simply reproducing each other's regulations (Weststrate & Weststrate 2009), international consistency of practice does imply a degree of consensus on the value of different forms of evidence. However, it has been observed (for example by Müller 2016) that there is considerable variation in the requirements for the nature and length of training or practice accepted as evidence for English language ability.

For registration in anglophone Canada (National Nursing Assessment Service, www.nnas.ca), all applicants whose first language is not English are required to take a language test. To qualify for a waiver, first language speakers of English must show that their nursing education program was taught in English (and not via an online or distance programme) and they must demonstrate that they have safely practiced nursing within the previous two years in a location

where the main language used was English and services were provided in English: i.e., completing a training programme is not considered sufficient.

The New Zealand Nursing Council (NZNC, www.nursingcouncil.org.nz) states that applicants for nursing registration (whether first language English speakers or not) may request a waiver of the language testing requirement if they have completed their nursing education and are registered as a nurse in the UK, Ireland, Canada or the USA. Different rules apply to applicants arriving from Australia under the Trans-Tasman Mutual Recognition Act. The Nursing and Midwifery Board of Ireland (NMBI, www.nmbi.ie) also recognises completion of nursing courses taught and assessed entirely in English in the UK, Australia, Canada, New Zealand, or the USA. Language tests are also waived for those who can demonstrate three years of post-registration practice within the previous five years in one of the recognised countries.

The Australian Health Practitioner Regulation Agency (AHPRA, www.ahpra.gov.au) and the Nursing and Midwifery Board of Australia (NMBA) waive language tests for nurses and midwives who have completed a nursing programme taught and assessed solely in English in a recognised country if, additionally their "primary language" is English, and they have completed six years of primary and secondary education taught and assessed solely in English. If their primary language is not English, they must have completed a total of five years of continuous education, including their professional qualification, through English. Unlike the NZNC, NMBI and NMC, the AHPRA includes South Africa on its list of recognised English-speaking countries.

In the US, State Boards of Nursing each set their own requirements (www.cgfns.org). Graduates of programmes taught in English in any country listed as English speaking are exempt from English language tests, but the lists of recognised countries differ by state. Among the more liberal, the Arizona State Board of nursing recognises Trinidad, Tobago, Ghana, Jamaica, and Barbados in addition to the countries recognised by the AHPRA. Many other states (such as New Hampshire or Tennessee) exclude Ghana from their lists, and some (such as Wisconsin) only recognise the same countries as the NZNC and NMBI. In Arizona, English test waivers are available for applicants employed as a nurse for 960 hours or more within the previous five years in one of the listed countries. This option does not appear to be available in most other states.

In short, international practice varies on the extent of educational experience or practice needed to evidence English language abilities. The NMC recognises more countries as English speaking than most other regulators but excludes countries such as Ghana and South Africa that are recognised elsewhere. The NMC requirement for at least a year of practice completed within the two years preceding the application is less restrictive than most other jurisdictions, but the justification for regulators determining that six months, a year, or a longer period of practice within two, or five years of application are not transparent. As noted by Clanfield (2019), evidence for the persistence of healthcare-relevant language abilities (which informs time limits on the validity of test results) is sparse.

While it can be argued that consultation on current policies has not revealed widespread concern about low levels of English abilities, this does not rule out the possibility that the use of lower test scores or of other forms of evidence of English language abilities might be accepted without risking safety or effectiveness. Indeed, recent consultations led to a relaxation rather

than a tightening of language standards (NMC 2018a). Critics of current policy have called for even greater flexibility. Pradeep and Davis (2022), for example, have commented that applicants who are unable to meet language test requirements, but that have worked in other roles in the NHS over an extended period (often as healthcare assistants) should be registered on that basis, arguing on domain description grounds that the language tests do not reflect the realities of clinical practice and that their use has discriminatory consequences. This has led to the launch of a public campaign (Das 2022) and a question raised in Council by Peter Mount on 26 January 2022 suggesting that the current policy "can be seen as racist, discriminatory and to support exploitation".

Arguments made by the campaigners that a period of residence in the UK or passing an English test for citizenship might be taken as evidence for adequate language abilities for safe clinical practice do not stand up to scrutiny. Tests of languages for citizenship may include very different content (limiting their domain relevance), and often require a lower level of proficiency than tests for professional registration. Migrants may live in a country for decades without ever acquiring the language abilities needed to support professional employment. However, we acknowledge the gap between testing and practice and the difficulties that this could cause for qualified individuals who struggle to evidence their true abilities through testing. We consider that opportunities for periods of supervised practice (with structured observation of language performance) for those scoring a little below the current thresholds might offer an alternative route to registration, paralleling the education and practice pathways.

When it comes to language tests, most nursing and midwifery regulators internationally accept IELTS and OET as evidence for English language abilities. In New Zealand (NZNC), Ireland (NMBI) and the UK, these are the only tests accepted. Alongside IELTS, CELBAN (but not OET) is recognised by provincial Colleges of Nurses and Midwives in Canada. The Australian NMBA recognises both IELTS and OET as well as the American TOEFL internet-based test (iBT), and the Pearson test of English (Academic) PTE-A. In the US, TOEFL-iBT is the default with most state boards also accepting IELTS. Florida and Oregon accept the OET. Many, including California, Illinois and Texas, accept the Michigan English Test and a few accept the Test of English for International Communication (TOEIC) (such as Arkansas and Louisiana) or the PTE-A (Arizona, Massachusetts).

Rules concerning the combination of test scores from different sittings and the validity of certificates also vary. The NMC will accept scores combined from more than one sitting if the tests are taken within six months of each other and scores on all sections or components of the test are not below a minimum score. All scores must be less than two years old at the time of application. Similar rules apply in Australia. The NMBA accepts scores from two sittings with no component below a stated minimum. The first test must be taken within the two years before the date of application. In New Zealand, the NZNC accepts scores from multiple sittings within 12 months if all scores are less than three years old. In Canada, Ireland and the USA, scores must come from a single sitting and must be less than two years old at the time of application.

Concerns about the equivalence of different pathways to registration became an issue of contention in Australia when an influential paper by Birrell (2006), an advisor to the Australian

immigration minister, reported that around a third of students graduating from Australian universities and applying for permanent residence failed to obtain the overall IELTS band score of 6.0 equated with 'competent English' for immigration purposes, even though this was equal to, or below the standard required for entry to most undergraduate programmes. Subsequent smaller scale but more controlled studies such as O'Loughlin and Arkoudis (2009) and Knoch, Rouhshad and Storch (2014) also found that students generally made little or no improvement in their English language test (IELTS) scores when retested at the end of their degree courses. The 63 students in the O'Loughlin and Arkoudis (2009) study, for example, made an average improvement of just 0.41 bands on the nine-band scale on a test administered at the end of their courses (3 years for the 30 undergraduates, 1 year for the 33 postgraduates) over the score they submitted for admission to university. The lowest levels of improvement were for writing: an average of just 0.21 bands.

A brief survey of entry requirements for international students on nursing programmes in the east of England (as stated on university websites) revealed that some require the same or higher IELTS scores than those required for NMC registration. The University of Essex matches the NMC by requiring a minimum IELTS score of 6.5 in writing and 7.0 in all other components; the University of East Anglia and the University of Northampton both require a minimum of 7.0 in all components. Others have rather lower entry requirements. Anglia Ruskin University, the University of Bedfordshire, and the University of Hertfordshire all require IELTS 6.0 overall with no less than 5.5 in any component. If improvements in proficiency (as measured by IELTS) were to be as low on these programmes as have been observed in Australia, it is possible that students with relatively poor English might be accepted for registration through this route. It might prove informative to relate the kinds of evidence of English language abilities used at registration to any subsequent communication difficulties of the kinds identified by (among many others) McKitterick et al. (2022).

Equally, it has been observed that not all first language speakers of English obtain high scores on IELTS or similar tests. As part of the original IELTS development process, a study by Hamilton et al. (1993) found that first language speakers educated to degree level averaged no better than 6.5 on the Writing test and 5.5 on Reading, although the researchers acknowledged that the participants' levels of motivation may not have been high. IELTS test data for 2019 (www.ielts.org/for-researchers/test-statistics/test-taker-performance) shows self-reported first language English speakers scoring an overall average of 6.9, with 6.7 for Reading, 7.2 for Listening, 6.2 for Writing and 7.1 for Speaking. There are reports that many first language English speaking healthcare professionals, including nurses, struggle to achieve scores on IELTS required for professional registration (McKew 2017). Occasionally, this has resulted in international *causes célèbres*, such as the case of Louise Kennedy, a qualified and experienced Irish vet who failed to meet the English language requirement when applying for an Australian settlement visa (Collins 2017).

4.1. Evaluation and Generalization as bases for test recognition

To accept the claims for an assessment, the NMC would need to establish that administration of the assessment follows rigorous guidelines; that procedures are in place to ensure compliance with these guidelines; that the storage and delivery of all test material is secure;

that robust procedures are in place to prevent and detect all forms of assessment malpractice; that appropriate action will be taken if malpractice is detected; and that results submitted to the NMC are verifiable.

Assurance should be sought that different forms of the assessment are constructed following a standard set of specifications; that material is scrutinized to ensure that it follows the specifications; that material is piloted with representative groups of test takers to ensure that it is of appropriate difficulty and quality for operational use; that there is content-based and statistical evidence for the equivalence of assessment forms; that scoring conforms to standard answer keys; that scores awarded through judgement are supported by training and scoring guides; that examiners are monitored and are consistent in the scores they award; that scores on each component of the assessment (especially scores at a level that affects success or failure) are acceptably reliable. Assessment providers should supply the NMC with estimates of conditional standard errors in each skill assessed for the score level that determines whether the applicant has satisfied the English language requirement.

4.2. Evaluation and generalization: conclusions

Our **key conclusions** from our review of the issue of evaluation and generalization can be summarised as follows:

As conditions for recognition of an assessment by the NMC, assessment providers should

provide

- evidence of their organisational status and governance
- evidence for the rigour of their administration, production, scoring, security and verification procedures
- documentation of the measurement properties of their assessments.

More evidence is needed for the parameters for recognising periods of study or practice as evidence for English language abilities.

Tracking data should be collected to monitor the relationships between different types of evidence of English language abilities and evidence of communicative success post-registration (such as reports of difficulties and follow-up studies)

Given the inevitable limitations of language tests, additional alternative pathways could be considered as routes to registration (such as periods of supervised practice). We note that supervised practice, combined with the overseas nursing programme and lower IELTS test score requirements) was formerly recognised as a route to registration and a variation on this could be introduced, incorporating a structured observation of English language performance.

5. Explanation

This step in the argument involves relating assessment results to a theory that accounts for performance.

Claim:

The theoretical construct informing assessment design is borne out by evidence from the results.

This claim guides the investigation of test results to establish whether they are consistent with the theoretical construct on which the assessment is based. It addresses the twin issues of construct irrelevance and construct representation. These are the extent to which performance on an assessment reflects the targeted language abilities (rather than other, unintended factors) and whether it addresses those abilities comprehensively. If the theory of English language abilities that informs the assessment design is sound, it should be possible, for example, to predict which tasks will prove relatively easy or difficult for test takers and which skills or strategies they will employ when they respond. The assessment results should be consistent with results of other assessments of the same or similar abilities, or with teachers' or colleagues' judgements of which language users have stronger or weaker abilities in English. Patterns of results within the assessment should reflect the elements of the theory of language on which the assessment is based (for example, if the theory predicts that 'reading for gist' is a distinct skill from 'reading for detail' results for items targeting each skill should follow distinct patterns. Depending on their profile of abilities, some test takers will perform relatively better on one item type than the other.

6. Explanation as a basis for test recognition

There are many ways of investigating whether explanation claims are justified, but the important issue for the NMC will be whether the assessment provider has a comprehensive programme or research and publication to support and disseminate research that investigates validity. Test providers like the IELTS partners and Cambridge Boxhill Language Assessment (for OET) conduct their own research and support studies by independent researchers into the qualities of their tests. Other providers should be able to demonstrate a similar commitment.

6.1. Explanation: conclusions

Our **key conclusions** from our review of the issue of explanation can be summarised as follows:

Assessment providers should provide evidence to support the validity of the theories on which their assessments are based. This could include an account of how the content and level of the assessment relates to an external framework such as the Common European Framework of Reference for Languages (Council of Europe 2001).

Assessment providers should demonstrate an ongoing commitment to validity research.

7. Extrapolation and Utilization

In this context, the key issue for the extrapolation step in the argument is whether test scores accurately predict the applicant's ability to understand and communicate through English in their nursing/ midwifery practice. Utilization can be understood in broader terms, but an essential question for the NMC is the appropriateness of the performance standards used to determine whether an applicant meets the English language requirement.

Extrapolation

Claim:

The estimate of an individual's English language abilities provided by the assessment provides useful information for determining whether they will have adequate English language abilities to support safe and effective practice.

Utilization

Claim:

Estimates of an applicant's English language abilities in healthcare contexts obtained from the assessment are useful for making decisions about registration.

Evidence supporting the extrapolation claim may come from the monitoring the performance of individuals who have been registered to confirm that they have the English language abilities needed for safe and effective practice. The relationship between the occurrence of complaints about IENMs' communication skills and the source of evidence provided by those IENMs at registration would be another valuable source of data. As noted above in the section on domain description, many studies that have explored the difficulties that IENMs may experience even after meeting the English language requirements. Adequately addressing such issues cannot realistically involve raising test score requirements but might involve educational programmes.

With respect to utilization, the standard psychometric approach to determining appropriate cut scores for professional registration is to conduct standard setting studies. The usual steps are as follows:

- First, define and assemble a group of experts
    - a group that combines knowledge of the subject, target context, assessment, and test taking population
    - recommendations on the minimum required size of the expert group vary, but the minimum is around 10
- Hold a workshop with the group of experts. Using one of a number of methods, ask them to estimate the score level that would represent a minimal acceptable level of ability for a defined purpose (such as safe and effective healthcare practice)
- Offer opportunities for discussion and provide feedback on matters such as the impact of setting different score requirements, item performance or test taker performance
- After discussion, allow the experts to revise their estimates. Combine the results to recommend an appropriate cut score for the intended purpose, taking account of variability in the experts' estimates.

In the USA, the National Council of State Boards of Nursing (NCSBN) has carried out a series of standard setting studies for English language tests including the Test of English as a Foreign Language (TOEFL) (O'Neill 2004, O'Neill, Marks, & Wendt 2005, O'Neill, Tannenbaum, & Tiffen 2005, Wendt, & Woo 2009, Wendt, Woo, & Kenny 2009), IELTS (O'Neill, Buckendahl, Plake, & Taylor 2007), the Michigan English Language Assessment Battery (MELAB) (Qian, Woo, & Banerjee 2014), and the Pearson Test of English: Academic (PTE-A) (Woo, Dickinson, & de Jong 2010). These studies involved the assessment providers, but were led by the NCSBN, guarding against suggestions of commercial influence. The outcomes are summarised in Table 1.

Table 1 NCSBN standard setting study score recommendations

TOEFL iBT	Wendt & Woo 2009	overall score of 84 with a minimum of 26 for speaking

IELTS	O'Neill et al. 2007	overall score of 6.5, no less than 6.0 in any module
Pearson Test of English: Academic	Woo et al. 2010	overall score of 55 with no subscores lower than 50
Michigan	Qian et al. 2014	overall score of 81 with a speaking score of 3

Table 2 Summary of score concordance tables presented by test providers

Provider	Concordance	Score levels									
ETS[1]	IELTS	6		6.5			7			7.5	
	TOEFL iBT	60		79			94			102	
IELTS[2]	IELTS	6		6.5			7			7.5	
	PTE-A (writing)	51.6 (62.2)		58.5 (74.1)			66.3 (82.3)			74.6 (87.5)	
Pearson[3]	PTE-A	46	53	56	59	64	66	68	72	76	78
	IELTS	6		6.5			7			7.5	
	TOEFL iBT	65-66	79-80		87-88	94	97	99-100	105		113
Michigan[4]	TOEFL iBT	61-67	76-82		83-91	92-99		100-104	105-110		111-116
	MELAB (MET)	69-72 (46-48)	76-77 (51-52)		78-81 (52-55)	82-83 (56-57)		84-87 (58-61)	87-91 (61-64)		92-93 (65-)

In arriving at their recommendations for the various tests, the results from the NCBN studies are reasonably consistent with concordance tables published by the various test providers (Table 2) that compare the level of English language ability represented by scores on each test (to the extent that the content of these tests is comparable). The recommended overall cut score for PTE-A of 55 appears to be marginally lower than the IELTS 6.5 level according to both the IELTS partners and Pearson while an overall score of 84 on TOEFL iBT would fall within the IELTS 6.5 band according to the table published by ETS.

Most state boards of nursing use the NCSBN recommendations for TOEFL, although a few set lower requirements (Florida, 76 overall, Michigan 80 overall). For IELTS, an overall minimum requirement of 6.5 is widely applied. Some states (such as Alabama, Massachusetts, New Jersey) follow the original NCSBN recommendations of no less than 6.0 on any module, but

[1] https://www.ets.org/toefl/score-users/scores-admissions/compare
[2] https://ielts.com.au/australia/about/article-which-ielts-test-convert-ielts-to-pte-test-score#how-should-your-organisation-use-the-2021-ielts-vs-pte-score-equivalencies
[3] https://www.pearsonpte.com/research/scoring
[4] https://michiganassessment.org/wp-content/uploads/2020/02/20.02.pdf.Res_.MELABComputerBasedTOEFLStudy.pdf

others (perhaps following the practice for TOEFL) ask for a higher score in Speaking: typically, 7.0 (as in Colorado, Maine, Wisconsin).

Outside the USA, the basis for determining cut scores is less transparent. In Australia and New Zealand, the NMBA and NZNC require IELTS 7.0 on all components, or OET B, with B (350) on each component. However, the Midwifery Council of New Zealand sets a higher requirement of 7.5 with no less than 7.0 on any component on IELTS but, like the NZNC, accepts an overall B with B (350) on each component for OET. Canadian provincial nursing boards require an overall 7.0 on IELTS with no less than 6.5 in writing and 7.5 in listening. The Irish NMBI mirrors the NMC in requiring 7.0 overall and on each IELTS component except writing, set at 6.5 and, for OET, the B overall and on each component except writing, set at C+ (300).

The NMC standards take this (shifting) international context into consideration as well as regular national consultations, and developments in healthcare practice, language assessment and training programmes. Perhaps because of this, the requirements are consistent with practices elsewhere: a little higher than in the USA, very similar to those in Canada, Australia and New Zealand and essentially the same as in Ireland. Although other regulators (unlike the NMC) have not set lower requirements for writing, there is a more general trend towards setting higher cut scores for spoken language skills (speaking in the USA, listening in Canada) and lower cut scores for the written language (writing in the UK and Ireland, reading in Canada).

Although consultation and international comparisons are important, they carry with them a risk that policy may be unduly affected by lobbying or by developments elsewhere. IENMs themselves tend to regard the tests as too difficult. Gillin and Smith (2021) for example found "There was also consensus amongst participants that the IELTS was unnecessarily difficult compared with what was expected and required of them in practice, especially in terms of writing. An Academic IELTS level 7 pass in writing was considered a disproportionate requirement for what sufficed in reality" (p.3). Employers recruiting nurses internationally may also be frustrated by failure rates on these tests. Conversely, Müller (2011) noted that according to the IELTS partners, "an IELTS test score of 7.0 – the professional registration requirement – is considered 'probably acceptable' to begin a health-related university degree." (A20). Regular standard setting exercises of the type employed by the NCSBN provide an evidence base that may help offset pressures from these sources. With this in mind, it would be prudent to adopt the NCSBN approach of directing standard setting efforts independent of the assessment providers.

The occurrence of cases where fitness to practice has been challenged on grounds of "not having the necessary knowledge of English" appears to be low. Although we have not located detailed statistics on this issue, we did find evidence of four registrants being struck off in 2021 for this reason (a total of 56 registrants were struck off in 2020-2021: NMC 2021). Summary data for April to September 2018 indicated that 13 of 289 registrants were struck off for reasons connected to "communication" (NMC 2018c). This suggests that the current NMC requirements are consistent with safe and effective practice but does not rule out the possibility that additional applicants with lower test scores, or who satisfied language

requirements through alternative pathways might also be capable of safe and effective practice.

Although pricing is not often included in arguments for assessment validity, a criticism made of the use of language tests for registration is that these (and associated preparation courses and materials) are very expensive, and that the costs are borne by the applicants. When addressing such criticisms, it might be helpful for the NMC to understand how prices are calculated and how they contribute to the qualities of the assessment.

7.1. Extrapolation and Utilization: conclusions

Our **key conclusions** from our review of the issue of extrapolation and utilization can be summarised as follows:

The NMC should

- systematically collect data and analyse the performance of IENMs gaining entry to the register via each of the available evidence pathways. It is important to identify how any registrant subsequently found to have a lack of knowledge of English satisfied the requirements on entry
- lead periodic standard setting studies (at intervals of 5 to 10 years or following substantial revisions to the relevant assessment) to strengthen the evidence base for cut scores on English language tests as a prerequisite for registration: Such studies could be led by the NMC but funded by assessment providers
- require test providers seeking recognition to explain how preparation for their tests will encourage language learners to develop language abilities appropriate to healthcare contexts
- require justifications for test providers' pricing policies

8. Conclusions

Here we briefly respond to the two questions we were asked to consider

Is the approach to language testing currently adopted by the NMC proportionate and appropriate?

The NMC's approach is founded on consultation, the practices of other healthcare regulators internationally and reviews of the research literature. These are all examples of good practice and there is little evidence to suggest that large numbers of nurses, midwives or nursing associates with inadequate English language abilities are being registered.

Although some forms of language assessment better represent the communication needs of nurses, midwives and nursing associates than others, it is widely acknowledged that no language test can do this comprehensively.

There is a risk that the discrepancies between test content and healthcare practice may give rise to disadvantage for certain groups. The extent of any such disadvantage should be investigated and mitigated.

There is also a clear implication that safety and effectiveness will benefit if further training and support for communication can be provided for registrants who have satisfied language requirements.

The rationale for the three forms of evidence currently accepted by the NMC is not sufficiently transparent. It should be possible to set out considerations and criteria for the recognition of tests and other forms of evidence that would allow providers of assessments other than those already granted recognition to make their case, to guide the development of new assessments or to inform the revision of existing ones in ways that would serve the NMC's purpose.

All assessments should be reviewed periodically to ensure that they remain fit for purpose, meeting the changing needs of nurses, midwives and nursing associates.

There may be scope for the recognition of alternative means for evidencing English language abilities such as the reintroduction of periods of supervised practice (with structured observation of their use of language) for professionals who score a little below the current requirements, but who meet other conditions including, for example, passing the test of competence and carrying out an extended period of service in a healthcare role such as working as a nursing assistant. This might compensate for difficulties encountered by professionals from some groups resulting from discrepancies noted above between testing and practice.

What would constitute an appropriate and practical methodology to investigate whether language tests of interest should be accepted by the NMC?

We recommend that the NMC should at regular intervals of three to five years convene a committee to consider proposals for assessments that would meet the NMC's requirements. These could include proposals for tests or for alternative sources of evidence. The onus would be on the assessment provider or sponsor to present evidence in support of their case for recognition. Recognition would be granted for a fixed period – probably until the next scheduled review. Improvements could be requested by the NMC with satisfactory implementation of these made a condition of continuing recognition.

A suitable framework for the recognition of language tests would build on the seven requirements currently listed in the NMC Code (NMC 2018b). We recognise some of these as compatible with language testing standards:
22.2 it tests reading, writing, listening and speaking;
22.4 there are high level security procedures for the test production and delivery;
22.5 there are test score verification systems that allow the NMC to confirm your results;
22.6 there is appropriate evidence of the reliability of the test and the dependability of its scores
But we suggest that others do not reflect widely accepted standards. Requirement *22.7 there are test centres widely available* appears outmoded while the rationales for *22.1 it tests knowledge of English in either a healthcare or academic context. It must not be a general test;* and *22.3 the speaking element is tested face-to-face and not via a computer test;* are unclear to us.

In the following section, we propose a framework that could be used to evaluate assessments that are put forward as suitable forms of evidence.

10. A framework for the recognition of language assessments by the NMC

We suggest the following as an outline framework to guide the evaluation of language tests by an advisory panel of experts convened by the NMC. With further refinements, this should provide panellists with sufficient information to judge whether a test meets industry standards and whether it should be recognised by the NMC. The framework, combined with feedback from panellists, would also serve to inform the development of suitable assessments by new providers and guide improvements to existing assessments so that, over time, these come to better serve the needs of the NMC. Continuing recognition could be made contingent on specified revisions being made before the next formal review.

The awarding organisation:

Provide evidence for the legal status and good governance of the assessment provider

Explain how the assessment provider's organisational structure supports the development, delivery and award of high-quality assessments

Domain definition:

Show how the proposed assessment tasks represent the English language abilities described in the NMC *Code.*

What aspects of communication in healthcare settings that are important to nursing and midwifery practice are not covered by the proposed assessment? How would you recommend that the NMC should address these aspects?

Evaluation:

Detail the management of the administration of the assessment: provide copies of relevant guidance for administrators

Demonstrate that the storage and delivery of all test material is secure

Describe the procedures for the prevention and detection of all forms of assessment malpractice

Describe actions taken if malpractice is detected

Describe procedures for the verification of results submitted to the NMC are verifiable.

Provide copies of the assessment specifications used in writing and assembling the assessment

Describe procedures for writing, reviewing, editing and evaluating assessment material

Describe procedures for piloting assessment material

Generalization:

Provide content-based and statistical evidence for the equivalence of assessment forms

Describe procedures for scoring performance and ensuring the consistency of scores

Describe how scorers are trained and monitored

Provide copies/ samples of training and scoring guides

Describe procedures for estimating the reliability of scores

Provide estimates of the assessment's conditional standard error in each skill tested at the (proposed) score level that determines whether the applicant has satisfied the English language requirement

Explanation:

Provide evidence to support the validity of the theory or theories on which the assessment is based

Provide evidence of a validation plan to support ongoing use of the assessment

Extrapolation:

Provide evidence that the assessment accurately predicts the ability of test takers to communicate safely and effectively in healthcare settings [note that this may require collaboration from the NMC to track applicants post-registration]

Utilization:

Provide evidence of the score levels on the assessment and its component parts that represent the minimum acceptable level of English required for safe and effective practice as a nurse, midwife or nursing associate in the UK

Explain how preparation for the assessment develops relevant communication skills for nurses, midwives or nursing associates.

Justify the pricing of the assessment in relation to the assessment provider's costs

Describe the support offered by the assessment provider to help language learners to develop relevant communication skills for nurses, midwives and nursing associates.

Provide samples of guidance given to test takers on effective means of preparing for the assessment

We gratefully acknowledge the input of Mark Henebury, ELT & Teacher Training Sector Specialist, Department for International Trade and Gad Lim, Head of Assessment and Research. Occupational English Test and of participants in the Nursing and Midwifery Council Workshop on 27 April 2022 who reviewed drafts of this report.

References

Al-Hamdan, Z. M., Al-Nawafleh, A. H., Bawadi, H. A., James, V., Matiti, M., & Hagerty, B. M. (2015). Experiencing transformation: the case of Jordanian nurse immigrating to the UK. *Journal of Clinical Nursing*, *24*(15-16), 2305-2313.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for Educational and Psychological Testing*. AERA.

Association of Language Testers in Europe. (1994). *ALTE code of practice*. ALTE.

Baldwin, S., & Cheng, L. (2020). Internationally Educated Nurses and the Canadian English Language Benchmark Assessment for Nurses: A Qualitative Test Validation Study of Test-Taker Accounts. *Canadian Journal of Applied Linguistics 23*(2), 96–117.

Blackmore, J., & Rahimi, M. (2019). How "Best Fit" Excludes International Graduates from Employment in Australia: A Bourdeusian Perspective. *Journal of Education and Work*, 32(5), 436–448.

Baude, D. (2022) English Language Timeline/Chronology. Unpublished briefing paper. NMC.

Birrell, B. (2006). Implications of low English standards among overseas students at Australian universities. *People and Place, 14*(4), 53-64.

Carr, A. (2021). OET vs. IELTS: Finding the most appropriate way to test language skills for medicine. *ESP Today 9*(1): 89-106.

Chapelle, C.A., Enright, M.K., and Jamieson, J.M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. Routledge.

Chok, H. N., Mannix, J., Dickson, C., & Wilkes, L. (2018). The factors impacting personal and professional experiences of migrant nurses in Australia: An integrative review. *Collegian, 25*(2), 247-253

Clanfield, L. (2019a). How appropriate and effective are different methods of testing language skills? Unpublished briefing paper. NMC.

Clanfield, L. (2019b). Language Testing using IELTS and OET – An Evidence Review. Update to 2018 review. Unpublished briefing paper. NMC.

Collins, P. (2017). Irish vet fails oral English test for Australian visa. *The Irish Times,* 8 August, 2017.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Das, S. (2022) Overseas nurses barred from profession due to 'discriminatory' language exam. *The Observer,* 30 April 2022.

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge University Press.

Ghazal, L. V., Ma, C., Djukic, M., & Squires, A. (2020). Transition-to-US practice experiences of internationally educated nurses: an integrative review. *Western Journal of Nursing Research*, *42*(5), 373-392.

Gillin, N., & Smith, D. (2021). Filipino nurses' perspectives of the clinical and language competency requirements for nursing registration in England: A qualitative exploration. *Nurse Education in Practice, 56,* 103223.

Hackett, K. (2021). NMC frauds: the 11 bogus registrants whose cases tricked the nursing regulator. *Nursing Standard,* 13 September 2021**.**

Hamilton, J., Lopes, M., McNamara, T., & Sheridan, E. (1993). Rating scales and native speaker performance on a communicatively oriented EAP test. *Language Testing*, *10*(3), 337-353.

Ho, Y.C., & Coady, M.R. (2018). English as a Second Language Nurses in the United States: Culture, Communication, and Needs for Continuing Education. *Studies in Continuing Education*, 40(2), 212–233.

International Language Testing Association. (2007). *ILTA guidelines for practice*. www.iltaonline.com/page/ILTAGuidelinesforPractice

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1-73.

Knoch, U., Elder, C., Woodward-Kron, R., & Manias, E. (2020). Capturing domain expert perspectives in devising a rating scale for a health specific writing test: How close can we get? *Assessing Writing*, *46.*

Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university?. *Assessing Writing*, *21*, 1-17.

Lin, Y., Ding, Y. and Li, J. (2018). A literature review of research exploring the experiences of overseas nurses in the United Kingdom (2002–2017) *Frontiers of Nursing, 5*(1), pp.17-30.

Lynch, T. (2016). Is English Language Background an Indicator of Success in the English Tests Required for Nursing Registration in Australia? *International Journal of Arts & Sciences, 9*(1), p. 539–569

Macqueen, S., Pill, J. and Knoch, U. (2016). Language test as boundary object: perspectives from test users in the healthcare domain. *Language Testing*. 33(2): 271-288.

Mckew, M., (2017). 'Common-sense' overhaul for UK language tests. *Nursing Standard 32*(9), 9.

McKitterick, D. J., Corsini, N., Peters, M. D., Chiarella, M., & Eckert, M. (2022). International nursing students' perceptions and experiences of transition to the nursing workforce–A cross-sectional survey. *Nurse Education in Practice*, *59*, 103303.

Müller, A. (2011). Addressing the English language needs of international nursing students. *Journal of Academic Language and Learning, 5*(2), A14-A22.

Müller, A. (2016). Language proficiency and nursing registration. *International Journal of Nursing Studies*, *54*(1), 132–140.

Müller, A., & Daller, M. (2019). Predicting international students' clinical and academic grades using two language tests (IELTS and C-test): A correlational research study. *Nurse Education Today*, *72*, 6-11.

Nursing and Midwifery Council (2015). Summary of findings from our consultation on English language requirements and process for registration with the NMC. NMC.

Nursing and Midwifery Council (2018a). English Language Requirements. NMC/18/102. NMC.

Nursing and Midwifery Council (2018b). *The Code: Professional standards of practice and behaviour for nurses and midwives. Updated to reflect the regulation of nursing associates.* NMC.

Nursing and Midwifery Council (2018c). Fitness to practise outcomes: April to September 2018. NMC.

Nursing and Midwifery Council (2018d). Future nurse: Standards of proficiency for registered nurses. NMC.

Nursing and Midwifery Council (2020). Guidance on registration language requirements. NMC.

Nursing and Midwifery Council (n.d.). Registration Appeals – English Language Requirements: Frequently Asked Questions (FAQs). NMC.

Nursing and Midwifery Council (2021a) The Nursing and Midwifery Council (Education, Registration and Registration Appeals) Rules 2004 (SI 2004/1767). Unofficial consolidated text Effective from 1 January 2021. NMC.

Nursing and Midwifery Council (2021b) Panel Member Refresher Training: English language appeals. NMC.

Nursing and Midwifery Council (2021c) NMC English Language Roundtable 18 November 2021. NMC.

O'Loughlin, K., & Arkoudis, S. (2009). Investigating IELTS exit score gains in higher education. *IELTS Research Reports, 10*, 95-180. IELTS.

Ofqual (2020). *Criteria for Recognition: May 2020.* Ofqual.

Okougha, M., & Tilki, M. (2010). Experience of overseas nurses: the potential for misunderstanding. *British Journal of Nursing*, *19*(2), 102-106.

O'Neill, T. R. (2004). The minimum proficiency in English for entry-level nurses: TOEFL®. NCLEX Psychometric Research Brief.

O'Neill, T. R. (2004). The minimum proficiency in English for entry-level nurses. Retrieved from www.ncsbn.org/pdfs/TOEFL_Research_Brief_vol_1.pdf.

O'Neill, T. R., Buckendahl, C. W., Plake, B. S., & Taylor, L. (2007). Recommending a nursing-specific passing standard for the IELTS examination. *Language Assessment Quarterly, 4*(4), 295-317.

O'Neill, T. R., Marks, C., & Wendt, A. (2005). Recommending a minimum English proficiency standard for entry-level nursing. *JONA's Healthcare Law, Ethics, and Regulation, 7*(2), 56-58.

O'Neill, T. R., Tannenbaum, R. J., & Tiffen, J. (2005). Recommending a minimum English proficiency standard for entry-level nursing. *Journal of Nursing Measurement*, *13*(2), 129-146.

Philip, S., Manias, E., & Woodward, K.R. (2015). Nursing educator perspectives of overseas qualified nurses' intercultural clinical communication: barriers, enablers and engagement strategies. *Journal of Clinical Nursing*, 24, 17-18.

Pilcher, N., & Richards, K. (2017). Challenging the power invested in the International English Language Testing System (IELTS): Why determining 'English' preparedness needs to be undertaken within the subject context. *Power and Education*, *9*(1), 3-17.

Qian, H., Woo, A., & Banerjee, J. (2014). Setting an English language proficiency passing standard for entry-level nursing practice using the Michigan English Language Assessment Battery. National Council of State Boards of Nursing – NCLEX Technical Brief.

Roughsedge, P. (2022) Revisiting Language Testing Methods: How appropriate and effective are different methods of testing language skills? Unpublished briefing paper. NMC.

Savvidou, V. (2022). Language Testing using IELTS and OET: An update of the 2019 evidence review. Unpublished briefing paper. NMC.

Seguis, B., & Lim, G.S. (2020). The writing that nurses do: Investigating changes to standards over time. *Assessing Writing*, *46.*

Staples, S., Laflair, G. T., & Egbert, J. (2017). Comparing Language Use in Oral Proficiency Interviews to Target Domains: Conversational, Academic, and Professional Discourse. *Modern Language Journal*, 101(1), 194–213.

Stubbs, F. (2017). Recruitment of nurses from India and their experiences of an overseas nurses program. *Nursing and Critical Care 22* (3), 176–183.

Toulmin, S. (1969). *The Uses of Argument*. Cambridge University Press.

Tse, J. (2018) Language Testing using IELTS and OET – An Evidence Review. Unpublished briefing paper. NMC.

Tuohy D (2019) Effective intercultural communication in nursing. *Nursing Standard, 34*(2), 45-50.

Wendt, A., & Woo, E. A. (2009). A Minimum English Proficiency Standard for The Test of English as a Foreign Language™ Internet-Based Test (TOEFL® iBT). NCLEX Psychometric Research Brief.

Wendt, A., Woo, A., & Kenny, L. (2009). Setting a passing standard for English proficiency on the Internet-based Test of English as a Foreign Language. *JONA's Healthcare Law, Ethics and Regulation*, *11*(3), 85-90.

West, E., Nayar, S., Taskila, T., & Al-Haboubi, M. (2016). A Review of the Literature on the Experiences of Black, Minority and Internationally Recruited Nurses and Midwives in the UK Healthcare system. Unpublished report to the Nursing & Midwifery Council]. https://www. nmc. org. uk/globalassets/sitedocuments/otherpublications/literature-review. pdf.

Weststrate, J. and Weststrate, M. (2009). Overseas nurses suggest improvements to registration process. *Kai Tiaki Nursing New Zealand*. 15(5):24-25.

Wickett, D., & McCutcheon, H. (2002). Issues of Qualification Assessment for Nurses in a Global Market. *Nurse Education Today*, 22(1), 44–59.

Woo, A., Wendt, A., & Liu, W. (2009). NCLEX pass rates: An investigation into the effect of lag time and retake attempts. JONA's Healthcare Law, Ethics and Regulation, 11(1), 23-26.

Woo, A., Dickinson, P., & de Jong, E. J. (2010). Setting an English language proficiency passing standard for entry-level nursing practice using the Pearson Test of English Academic. NCLEX Technical Brief.