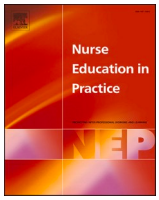
Nurse Education in Practice 81 (2024) 104199

Contents lists available at ScienceDirect

Nurse Education in Practice

journal homepage: www.elsevier.com/locate/issn/14715953



Development of qualified items for nursing education assessment: The progress testing experience

Bruna Moreno Dias [a,*,1,2], Lúcia Marta Giunta da Silva [b,3], Pedro Tadao Hamamoto Filho [c,4,5], Valdes Roberto Bollela [d,6], Carmen Silvia Gabriel [a,7]

[a] University of São Paulo, Ribeirão Preto College of Nursing, São Paulo, Brazil
[b] Federal University of São Paulo, Paulista School of Nursing, São Paulo, Brazil
[c] Universidade Estadual Paulista, Botucatu Medical School, Botucatu, Brazil
[d] University of São Paulo, Ribeirão Preto Medical School, São Paulo, Brazil

ARTICLE INFO

Keywords:
Education
Nursing
Baccalaureate
Nursing students
Academic performance
Formative feedback
Nursing faculty
Universities

ABSTRACT

Aim: To analyze the psychometric characteristics of items in the nursing inter-institutional progress testing for the years 2019, 2021, 2022 and 2023.
Background: Progress testing is a validated method for evaluating professional undergraduate education, aimed at identifying knowledge gain in a continuous and progressive manner, with potential benefits for nursing education. However, for its results to be useful, the evaluation items used in the test must have good psychometric performance.
Design: A cross-sectional study.
Methods: A sample of 377 items (multiple-choice questions) was applied to 4678 students in four years of progress testing. The difficulty and discrimination indexes were analyzed using descriptive statistics, ANOVA and simple linear regression.
Results: The average difficulty index of the test items ranged between 0.39 and 0.46. The areas of child and adolescent health, women's health and adult health had the most difficult items, while the areas of management, mental health and public health had the least difficult items. Discrimination index ranged from 0.35 to 0.43. There was a difference between discrimination index between the years of application ($p < 0.001$), with a significant increase in the discrimination index ($p < 0.001$) in the trend analysis. Students in the final years showed lower levels of difficulty and discrimination when compared with students in the initial years, demonstrating that the test is easier and there is less dispersion of performance among students in the final years.
Conclusions: The items are not difficult and have good discrimination. A gradual annual increase in the discrimination index of the items was observed. This study provides useful information for the psychometric analysis and quality assurance of knowledge assessment items, both for the implementation of similar PT experiences and in the use of multiple-choice questions for other knowledge assessment purposes.

* Corresponding author.
E-mail addresses: brunamorenodias@gmail.com (B. Moreno Dias), lsilva@unifesp.br (L.M. Giunta da Silva), pedro.hamamoto@unesp.br (P.T. Hamamoto Filho), vbollela@fmrp.usp.br (V.R. Bollela), cgabriel@eerp.usp.br (C.S. Gabriel).
[1] https://orcid.org/0000-0002-7346-4848.
[2] Twitter: @brunamdias.
[3] https://orcid.org/0000-0002-7737-0443.
[4] https://orcid.org/0000-0001-6436-9307.
[5] Twitter: @hamamoto_pedro.
[6] https://orcid.org/0000-0002-8221-4701.
[7] https://orcid.org/0000-0003-2666-2849.

https://doi.org/10.1016/j.nepr.2024.104199
Received 22 May 2024; Received in revised form 22 October 2024; Accepted 5 November 2024
Available online 7 November 2024

1. Introduction

Investing in nursing education is a strategic effort at a global level and it is necessary to optimize the educational system to align it with health needs and guarantee the provision of teaching programs with quality standards (World Health Organization, 2021). Given the growing concern about the quality of nurse education, evaluating the teaching-learning process in undergraduate nursing courses is essential (Fehn et al., 2021).

Student assessment should be based on cognitive and behavioral competencies (Miller, 1990), both important and complementary, with knowledge being the basis for professional performance (Lindgren et al., 2024). In this process, evaluations are expected to be summative, to ensure that minimum standards are met and formative, to improve the learning process. This makes it possible to recognize the student's ability to apply knowledge in practice and to identify and correct learning gaps (Villela et al., 2022).

Assessments are necessary to determine learning, retention and application of knowledge and for this, valid and reliable tools are needed (Neeley et al., 2016). One of the assessment tools in this context is progress testing, with comprehensive coverage of all the knowledge areas in the curriculum (Lindgren et al., 2024).

Progress testing (PT) is a cognitive assessment, aimed at verifying the student's gain in knowledge in a continuous and progressive manner (Bicudo et al., 2019). PT adds to the programmatic assessment instruments. Although it fulfills the summative purpose of assessment, it makes a special contribution to the formative purpose, with continuous, comprehensive assessments that do not focus exclusively on knowledge, but also on the ability to apply knowledge (Troncon et al., 2023).

The PT differs from other multiple-choice exams, applied in course evaluations or professional licensing exams, due to its longitudinal aspect, applied annually, with a formative assessment perspective, where feedback and possibilities are provided for the student to develop in subjects where he/she has not yet reached required competencies and does not have a classification or approval nature (Dias et al., 2024).

In the medical field, PT has a long history of application in different countries, such as the Netherlands (Tio et al., 2016), Canada (Blake et al., 1996), United Kingdom (Freeman et al., 2010), Germany (Görlich and Friederichs, 2021), Saudi Arabia (Alamro et al., 2023) and Brazil (Cecilio-Fernandes et al., 2021). In addition, there are reports of PT being applied to dental hygiene, dental therapy and dentistry students in the UK (K. Ali et al., 2018) and the dentistry course in Austria (Kirnbauer et al., 2018). However, there are no reports in the literature of the application of PT in nursing courses.

The subject of this study, the nursing PT, carried out by eight Brazilian public teaching institutions, is an innovative strategy for evaluating the education of nurses. Its content is based on the National Curriculum Guidelines for Undergraduate Nursing Programs, which establish the nurse's training profile, aiming to develop a qualified professional for the practice of nursing (Brasil. Ministério da Educação, 2001). It is important to mention that in Brazil there are no competency or licensing exams for newly graduated nurses. In this sense, the PT is an opportunity to evaluate and seek improvements in nursing education.

2. Background

PT allows comprehensive analysis of students' knowledge in different areas, enabling cross-sectional analysis, comparing performance in different skill groups and longitudinal analysis, comparing knowledge over time (Freeman et al., 2010).

The same test is administered simultaneously to all students, regardless of their year of study, in other words, students in the initial and final years take the same test at the same time. The test prioritizes the application of knowledge rather than simple memorization (Alamro et al., 2023) and its content focuses on the level of knowledge expected of a newly graduated professional (Hamamoto Filho et al., 2023).

Although it is recognized that PT should not be the only way of assessing knowledge, the analysis of its results is an important feedback tool for students and educational institutions (Cecilio-Fernandes et al., 2021). The progress testing provides feedback for all those involved in the educational process (Coombes et al., 2010). Students can situate themselves in the evolving learning process (Bicudo et al., 2019). In addition, they are able to self-assess skills milestones, identify knowledge gaps (Neeley et al., 2016), besides acquiring knowledge and developing the ability to apply and reflect on the content covered (Cecilio-Fernandes et al., 2016). On the other hand, educational institutions obtain an overview of the curriculum structure, for the specific evaluation of subjects or areas of knowledge or for the evaluation of curriculum amendments (Bicudo et al., 2019).

However, challenges include the time required to develop and implement PT and the difficulty of mapping items for the test in line with the required content and curriculum objectives (Neeley et al., 2016). In this sense, the blueprint is essential for constructing the test and ensuring its validity. Blueprint ensures coherence between educational objectives, learning and assessment contexts (McLaughlin et al., 2005). The blueprint makes it possible to define the purpose and scope of the test, the areas of knowledge and skills assessed. It is relevant for defining the composition of the test, the areas covered and the number of questions for each domain; so that the assessment is aligned with the objectives of the educational program and with adequate representation of relevant topics for education (Abdellatif et al., 2024; Green and Heales, 2023).

Student knowledge assessments commonly use multiple choice questions (Gottlieb et al., 2023). Questions may require lower or higher levels of cognitive processing, depending on whether students have to remember, minimally understand or apply their knowledge (Hamamoto Filho et al., 2020). In order to ensure that the questions accurately assess the students and provide meaningful data, it is important to adopt good practices when designing multiple-choice questions (Gottlieb et al., 2023).

Preparing questions is a challenging process in terms of capacities and development time for the faculty member, which requires initial orientation and continuous capacity building on writing multiple-choice questions (Green and Heales, 2023). In this way, the application of PT benefits from the operationalization of a process that takes into account faculty training, questions revision by a review committee, quality questions analysis, student performance and the psychometric functioning of the questions (Hamamoto Filho and Bicudo, 2020).

Psychometric properties of the PT focus on attributes such as validity, reliability, difficulty and discrimination. The analysis of post-test psychometric indicators is useful for evaluating the test and providing feedback to experts (Abdellatif et al., 2024), its interpretation produces relevant information both for revising and improving the questions and for improving teaching (McGahee and Ball, 2009).

This study aims to analyze the psychometric characteristics of items in the nursing inter-institutional progress testing for the years 2019, 2021, 2022 and 2023.

3. Methods

3.1. Study design

Cross-sectional study to analyze psychometric characteristics of items from the nursing inter-institutional PT. For the preparation of the study and the presentation of reports, the *Strengthening the Reporting of Observational Studies in Epidemiology* (STROBE) tool was used (Malta et al., 2010).

3.2. Setting

The PT has been developed by eight Brazilian public institutions: Ribeirão Preto College of Nursing (USP), School of Nursing (USP),

Paulista School of Nursing (Unifesp), Marília Medical School (FAMEMA), São José do Rio Preto Medical School (FAMERP), Botucatu Medical School (Unesp), School of Nursing (Unicamp) and the Federal University of São Carlos (UFSCar) (Dias et al., 2024).

The test is administered once a year, at the beginning of the fourth quarter of the academic year. The test is held under exam conditions, simultaneously for all students, from all undergraduate years at all institutions and lasts a maximum of four hours. Participation is voluntary and there is no cost to the student.

The items are prepared by faculty members from the participating institutions, who are trained annually in standards and good practices on item writing and exam preparation. The test topics are selected using a blueprint based on the National Curriculum Guidelines for nursing education (Brasil. Ministério da Educação, 2001). This blueprint was developed in a participatory process, with representatives from all the institutions in the consortium, experts in nursing education, with validation of the items in a panel of experts in each area, made up of representatives from all the institutions.

The test consists of 120 items (multiple-choice questions with single best answer) distributed in six areas: Public Health (n = 20), Management (n = 15), Child and Adolescent Health (n = 20), Women's Health (n = 20), Adult Health (n = 35) and Mental Health (n = 10). More than knowledge retention, the items seek to assess the application of knowledge in clinical and managerial situations. The items are validated by an inter-institutional panel of experts in each area and in nursing education. Since items are evaluated using their psychometric performance and the need for robust samples for this evaluation, the items are not tested prior the application. However, every year around 35 pre-tested questions with good psychometric performance in last years are used in the exam and 85 new questions are developed.

3.3. Sample

The sample consisted of all the 377 new items developed for PT, 120 in 2019, 86 in 2021, 85 in 2022 and 86 in 2023. Thus, of the 480 items used in the four PT applications, 103 pre-tested items were not considered in this analysis. For the psychometric analysis, all students who took the test were considered, excluding those whose performance was equal to or less than 25 % of the total score, either due to incompleteness or randomly recorded answers. In total the test was taken by 1105 students in 2019, 1138 in 2021, 1175 in 2022 and 1260 in 2023.

3.4. Data collection methods

The study used secondary data retrieved from the test application database. The psychometric characteristics of each item were analyzed using an Classical Test Theory approach (Sakai et al., 2011).

The difficulty index was calculated by the proportion of students who answered incorrectly each item, ranging from 0 to 1, where the closer to 1, the more difficult the question. It is classified as: very easy (below 0.19), easy (0.20–0.39), medium (0.40–0.79), difficult (0.80–0.89) and very difficult (above 0.90) (McGahee and Ball, 2009).

The discrimination index was calculated as the difference in correct answers for each item between the 27 % of students who performed well in the test and the 27 % who performed poorly (Kelley, 1939), ranging from − 1 to 1, where the closer to 1, the better the discrimination. It is classified as: 'poor item, should be rejected' (below 0.19), 'poor item, subject to reworking' (0.20–0.29), 'good, but subject to improvement' (0.30–0.39) and 'good' (above 0.40) (McGahee and Ball, 2009).

3.5. Data analysis

To analyze the difficulty and discrimination indices, the students were categorized by undergraduate year and grouped into two categories: initial years (compromising students in their 1st, 2nd and 3rd study years) and final years (encompassing students in their 4th and 5th study years).

Descriptive statistics were used for univariate and bivariate analysis of the variables under study. The ANOVA test was used, with Tukey's post-test and simple linear regression. A 5 % significance level and 95 % confidence interval were assumed for all analyses. IBM SPSS Statistics software, version 28 (IBM Corp., Armonk, N.Y., USA)., was used.

3.6. Ethical approval

The principles and recommendations for research involving human subjects were respected. All the participating institutions consented to the study. Since we dealt with secondary data and no student was identified, individual consent was not necessary. The project was reviewed and approved by the Research Ethics Committee, under opinion 5.865.717.

4. Results

In the analysis of difficulty, most of the items were categorized as medium difficulty, ranging from 43.5 % to 53.5 % of the total number of items in each test. There was a reduction in the number of difficult and very difficult items, which amounted to 6.6 % of the items in 2019 and in 2023 none of the items were classified as difficult or very difficult.

The average difficulty index of the test ranged between 0.39 and 0.46 (Fig. 1). The difficulty index was higher among students in the initial years, ranging from 0.42 to 0.48, while among students in the final years the index ranged from 0.30 to 0.40, indicating that the test is easier for students in the final years.

The areas of child and adolescent health, women's health and adult health had the most difficult items, while the areas of management, mental health and public health had the least difficult items (Fig. 2).

The categorization of item discrimination showed a progressive increase in items with better discrimination with each test; items classified as 'good but subject to improvement' and 'good' accounted for 68.33 % of all test items in 2019 and 87.2 % in 2023. The average discrimination index has increased year on year (Fig. 3), from 0.35 in 2019 to 0.43 in 2023.

Discrimination index was higher among students in the initial years, with a range between 0.35 and 0.43, when compared with the discrimination index among students in the final years, with a range between 0.31 and 0.37, demonstrating a homogeneous group of students in the final years, with less dispersion of their performance in the PT.

In general, there was a progressive improvement in the item discrimination index for all areas (Fig. 4). In the last year of application all areas, except management, had a discrimination index higher than 0.40. Average discrimination equal to or greater than 0.40 was observed in mental health in all four applications of the test; the highest average discrimination was observed in this area in the last application of the test (0.58 in 2023).

The difficult index showed no significant difference between the years of application ($p = 0.051$), although a significant difference was observed between the years 2019 and 2022 ($p = 0.044$).

Regarding the discrimination index, there was a difference between the years of application ($p < 0.001$), with significant differences between the years 2019 and 2022 ($p = 0.017$) and 2019 and 2023 ($p < 0.001$).

In the trend analysis of the psychometric indicators, there was a significant increase in the discrimination index ($F = 10.257$, $p < 0.001$, adj $R2 = 0.052$) (Table 1).

5. Discussion

The analysis of items from 2019 to 2023 indicated tests of medium difficulty and good discrimination index. In general, the test shows a disproportionate distribution of item difficulty, with a predominance of medium- and low-difficulty items when compared with higher-difficulty

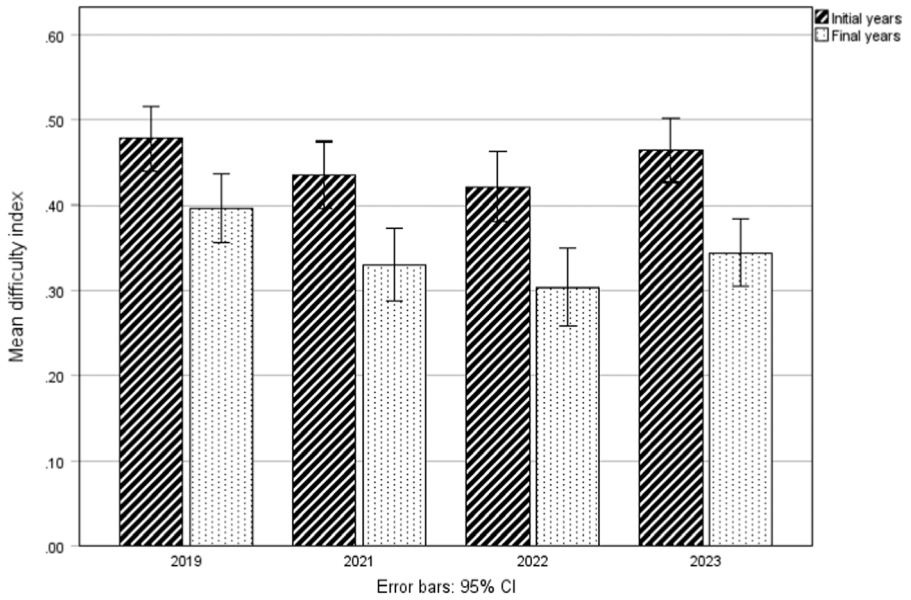


Fig. 1. Mean values of the difficulty index of the items applied in the progress testing in 2019, 2021, 2022 and 2023 by students' group.

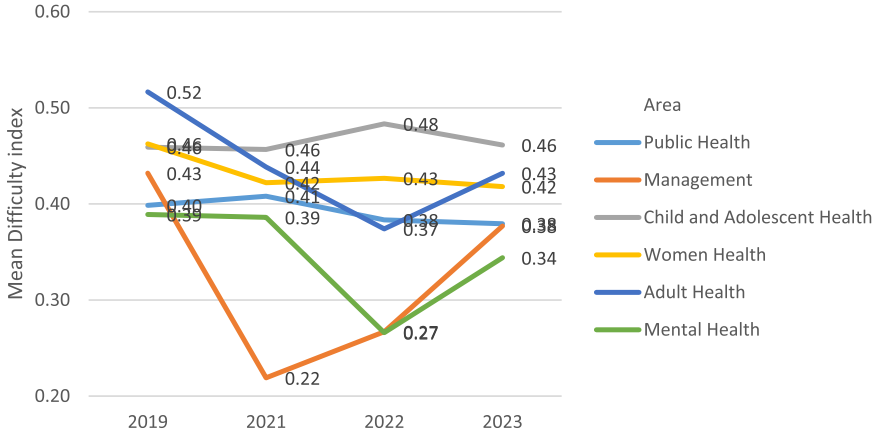


Fig. 2. Mean values of the difficulty index of the items applied in the progress testing in 2019, 2021, 2022 and 2023 by area.

items. On the other hand, there has been a consistent annual increase in the discrimination index with a greater frequency of good quality items.

The psychometric analysis of the items allows to identify the specific behavior of each area in the progress testing. In the area of management, although the items are of low difficulty, there has been an increase in the discrimination index. However, the public health items have low difficulty and lower discrimination index when compared with other areas, indicating an opportunity for improvement in the development of items in this area.

The difficulty index has remained medium and low (below 0.4), with stability in difficulty over the years. The predominance of medium and easy difficulty questions has been observed in other studies, such as the medical course, where these questions accounted for 90–95 % of the progress testing items (Feitosa Queiroz et al., 2022); and the experience of applying multiple-choice questions in dentistry course, where the easy questions varied between 45.8 % and 65 % of the items (Shaikh et al., 2020).

The proportion of good discrimination items, which in 2023 represented 87.2 % of the items, showed good performance when compared with the 60 % rate of good discrimination items in a progress testing of medicine (Feitosa Queiroz et al., 2022).

Despite the satisfactory results of the psychometric performance of the items, further analysis is needed to identify the technical quality of the test, with a view to improving the item development process and providing support for the development of faculty members who prepare questions (Villela et al., 2022).

Developing better quality questions increases the validity of the test (Shaikh et al., 2020). Faculty need to be familiar with the best practices for writing multiple-choice questions (Gupta et al., 2021). In this sense, providing faculty development on item writing questions is required and effective in reducing flaws (Gupta et al., 2020), because flawed items may threaten the examination's reliability (Ali and Ruit, 2015; Downing, 2005). However, some faculty members have little knowledge of how to identify flaws in the preparation of items (Kowash et al., 2020).

In this experience, a workshop is offered annually for faculty members on good practices for preparing multiple-choice questions. However, in addition to faculty development, other institutional actions favor the development of good items, such as working groups and panel review of the items (Bollela et al., 2018), improvements in communication and motivation for greater faculty engagement, the allocation of time for item development and the strengthening of peer review and feedback processes for item developers (Abdulghani et al., 2015, 2017; Karthikeyan et al., 2020).

The experiences of applying PT have implications and potential benefits for students, in assessing and supporting the learning process (Neeley et al., 2016); as a complementary method of assessing student

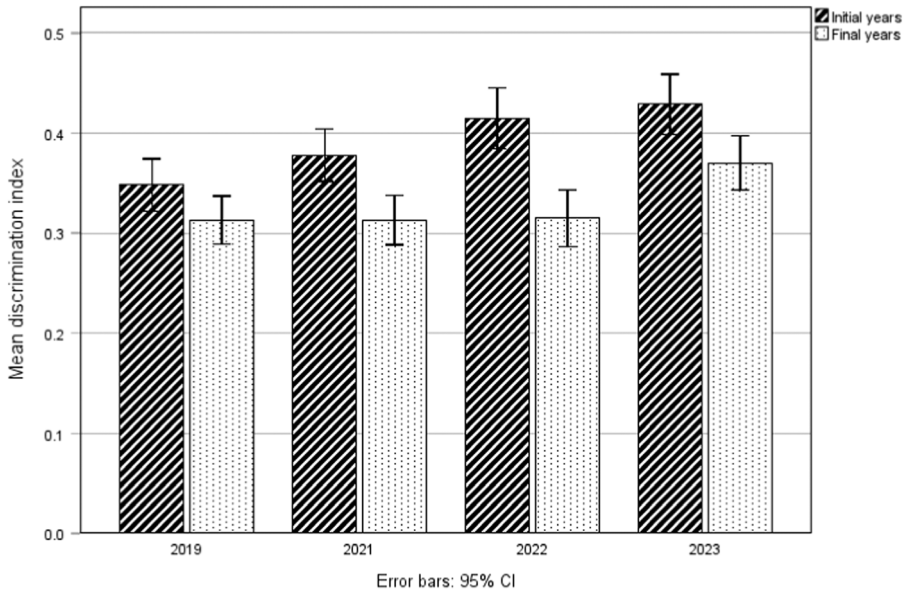


Fig. 3. Mean values of the discrimination index of the items applied in the progress testing in 2019, 2021, 2022 and 2023 by group.

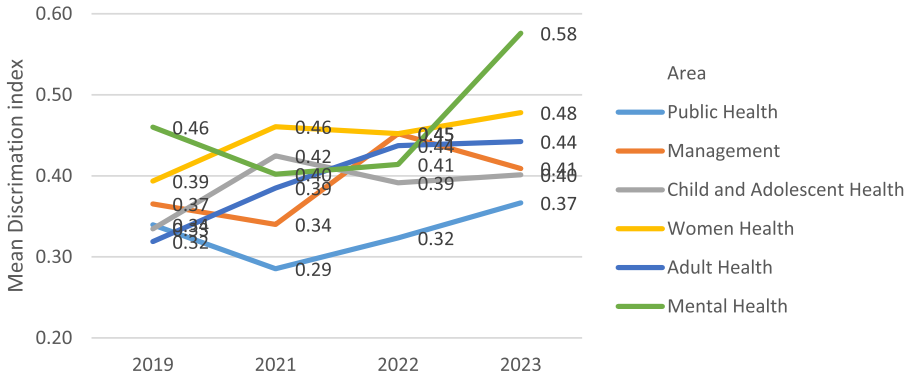


Fig. 4. Mean values of the discrimination index of the items applied in the progress testing in 2019, 2021, 2022 and 2023 by area.

Table 1
Results of the linear regression for trend behavior of psychometric indicators in the years 2019, 2021, 2022 and 2023.

Indicator	Standardized coefficients beta	P
Difficulty index	− 0.074	0.150
Discrimination index	0.202	< 0.001

performance by faculty members (Cecilio-Fernandes et al., 2021); and educational management, analyzing the effectiveness of the curriculum, identifying possible revisions and updates (Troncon et al., 2023). In the experience of applying the Nursing PT, the establishment of the consortium has allowed the test to be applied inter-institutionally, expanding the scope and quality of the test, with better accuracy in assessing the performance of nursing students (Dias et al., 2024).

6. Limitations and recommendations

A limitation of this study is that it was based only on psychometric characteristics, without evaluating the construct validity of the items and the suitability of the topic for the blueprint. Another useful point to be addressed in future studies is the role of low or high-cognitive order items on their psychometric behavior: it is probable that items of higher taxonomy, that require more complex cognitive skills, have better discrimination indices (Cecilio-Fernandes et al., 2018; Hamamoto Filho et al., 2020). Despite this, it is understood that this study provides useful information on the development of knowledge assessment items for faculty, educational institutions and researchers.

The findings of this study are especially useful for analyzing learning assessment processes, for implementing progress testing or for improving multiple-choice questions applied in other types of knowledge assessment. However, further studies could be developed in terms of the adoption of other statistical methods or even the joint analysis of psychometric data with information on the level of complexity of the cognitive assessment or the suitability of the questions for the blueprint and recommended structure.

7. Conclusions

The progress test has been demonstrated to be a useful tool for evaluating the retention and application of knowledge among undergraduate nursing students. The psychometric analysis of the characteristics of the items in the nursing inter-institutional progress testing indicate that the items are not difficult and have good discrimination.

A gradual annual increase in the discrimination index of the items was observed, due to efforts to improve the process of developing and reviewing the questions. Nevertheless, for future applications, the test should consider the development of questions of different levels of difficulty, with a higher proportion of questions of greater difficulty. In this

sense, in order to improve the quality of the items, actions aimed at capacity-building for faculty members in the development of multiple-choice questions are particularly useful.

The findings of this study provide useful information for the psychometric analysis and quality assurance of items, both for the implementation of similar PT experiences and in the use of multiple-choice questions for other knowledge assessment purposes.

CRediT authorship contribution statement

Pedro Tadao Hamamoto Filho: Writing – review & editing, Validation. **Lúcia Marta Giunta da Silva:** Writing – review & editing, Validation. **BRUNA MORENO DIAS:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Carmen Silvia Gabriel:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Valdes Roberto Bollela:** Writing – review & editing, Validation.

Registration number

[##].

Funding sources

Research grant awarded by the São Paulo State Research Foundation (FAPESP), process 2023/00554-8.

Conflict of interest

None declared by the authors.

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.

Acknowledgements

The authors are grateful for the collaboration of the faculty members of the Consortium for the Progress Testing of São Paulo Nursing Schools and all the students who took part in the assessments.

References

Abdellatif, H., Alsemeh, A.E., Khamis, T., Boulassel, M.-R., 2024. Exam blueprinting as a tool to overcome principal validity threats: a scoping review. Educ. Méd. 25 (3), 100906. https://doi.org/10.1016/j.edumed.2024.100906.

Abdulghani, H.M., Ahmad, F., Irshad, M., Khalil, M.S., Al-Shaikh, G.K., Syed, S., Aldrees, A.A., Alrowais, N., Haque, S., 2015. Faculty development programs improve the quality of Multiple Choice Questions items' writing. Sci. Rep. 5 (1), 9556. https://doi.org/10.1038/srep09556.

Abdulghani, H.M., Irshad, M., Haque, S., Ahmad, T., Sattar, K., Khalil, M.S., 2017. Effectiveness of longitudinal faculty development programs on MCQs items writing skills: a follow-up study. PLoS One 12 (10), e0185895. https://doi.org/10.1371/journal.pone.0185895.

Alamro, A.S., Alghasham, A.A., Al-Shobaili, H.A., Alhomaidan, H.T., Salem, T.A., Wadi, M.M., Saleh, M.N., 2023. 10 years of experience in adopting, implementing and evaluating progress testing for Saudi medical students. J. Taibah Univ. Med. Sci. 18 (1), 175–185. https://doi.org/10.1016/j.jtumed.2022.07.008.

Ali, K., Cockerill, J., Zahra, D., Tredwin, C., Ferguson, C., 2018. Impact of progress testing on the learning experiences of students in medicine, dentistry and dental therapy. BMC Med. Educ. 18 (1), 253. https://doi.org/10.1186/s12909-018-1357-1.

Ali, S.H., Ruit, K.G., 2015. The Impact of item flaws, testing at low cognitive level and low distractor functioning on multiple-choice question quality. Perspect. Med. Educ. 4 (5), 244–251. https://doi.org/10.1007/S40037-015-0212-X.

Bicudo, A.M., Hamamoto Filho, P.T., Abbade, J.F., Hafner, M. de L.M.B., Maffei, C.M.L., 2019. Teste de Progresso em Consórcios para Todas as Escolas Médicas do Brasil. Rev. Bras. Educ. Méd. 43 (4), 151–156. https://doi.org/10.1590/1981-52712015v43n4rb20190018.

Blake, J.M., Norman, G.R., Keane, D.R., Mueller, C.B., Cunnington, J., Didyk, N., 1996. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. Acad. Med. 71 (9), 1002–1007. https://doi.org/10.1097/00001888-199609000-00016.

Bollela, V.R., Borges, M. de C., Troncon, L.E. de A., 2018. Avaliação Somativa de Habilidades Cognitivas: Experiência Envolvendo Boas Práticas para a Elaboração de Testes de Múltipla Escolha e a Composição de Exames. Rev. Bras. Educ. Méd. 42 (4), 74–85. https://doi.org/10.1590/1981-52712015v42n4rb20160065.

Brasil. Ministério da Educação, 2001. Resolução CNE/CES nº. 3, de 7/11/2001. Institui Diretrizes curriculares nacionais do curso de graduação em enfermagem. DOU.

Cecilio-Fernandes, D., Kerdijk, W., Jaarsma, A.D. (Debbie) C., Tio, R.A., 2016. Development of cognitive processing and judgments of knowledge in medical students: analysis of progress test results. Med. Teach. 38 (11), 1125–1129. https://doi.org/10.3109/0142159X.2016.1170781.

Cecilio-Fernandes, D., Kerdijk, W., Bremers, A.J., Aalders, W., Tio, R.A., 2018. Comparison of the level of cognitive processing between case-based items and non-case-based items on the Interuniversity Progress Test of Medicine in the Netherlands. J. Educ. Eval. Health Prof. 15, 28. https://doi.org/10.3352/jeehp.2018.15.28.

Cecilio-Fernandes, D., Bicudo, A.M., Hamamoto Filho, P.T., 2021. Progress testing as a pattern of excellence for the assessment of medical students' knowledge: concepts, history and perspective. Medicina 54 (1), e173770. https://doi.org/10.11606/issn.2176-7262.rmrp.2021.173770.

Coombes, L., Ricketts, C., Freeman, A., Stratford, J., 2010. Beyond assessment: feedback for individuals and institutions based on the progress test. Med. Teach. 32 (6), 486–490. https://doi.org/10.3109/0142159X.2010.485652.

Dias, B.M., Silva, L.M.G. da, Salvetti, M. de G., Toledo, V.P., Tonhom, S.F. da R., Duarte, M.T.C., Irigoyen, B.B.T.J., Protti-Zanatta, S.T., Gabriel, C.S., 2024. Implementation of the São Paulo Nursing Courses Consortium for the Progress Test: experience report. Rev. Esc. Enferm. USP 58. https://doi.org/10.1590/1980-220x-reeusp-2023-0347en.

Downing, S.M., 2005. The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. Adv. Health Sci. Educ. 10 (2), 133–143. https://doi.org/10.1007/s10459-004-4019-5.

Fehn, A.C., Alves, T. dos S.G., Poz, M.R.D., 2021. Higher education privatization in Nursing in Brazil: profile, challenges and trends. Rev. Lat.-Am. Enferm. 29. https://doi.org/10.1590/1518-8345.4725.3417.

Feitosa Queiroz, É., Olivia Andréa Alencar Costa Bessa, iD., Daniela Chiesa, iD., 2022. Desempenho cognitivo dos estudantes de Medicina no Teste de Progresso. Rev. Bras. Educ. Méd. 46 (1). https://doi.org/10.1590/1981-5271V46.SUPL.1-20220305.

Freeman, A., Van Der Vleuten, C., Nouns, Z., Ricketts, C., 2010. Progress testing internationally. Med. Teach. 32 (6), 451–455. https://doi.org/10.3109/0142159X.2010.485231.

Görlich, D., Friederichs, H., 2021. Using longitudinal progress test data to determine the effect size of learning in undergraduate medical education – a retrospective, single-center, mixed model analysis of progress testing results. Med. Educ. Online 26 (1). https://doi.org/10.1080/10872981.2021.1972505.

Gottlieb, M., Bailitz, J., Fix, M., Shappell, E., Wagner, M.J., 2023. Educator's blueprint: a how-to guide for developing high-quality multiple-choice questions. AEM Educ. Train. 7 (1). https://doi.org/10.1002/aet2.10836.

Green, D.J., Heales, C.J., 2023. Progress testing: an educational perspective exploring the rationale for progress testing and its introduction into a Diagnostic Radiography curriculum. J. Med. Imaging Radiat. Sci. 54 (1), 35–42. https://doi.org/10.1016/j.jmir.2022.12.009.

Gupta, P., Meena, P., Khan, A., Malhotra, R., Singh, T., 2020. Effect of faculty training on quality of multiple-choice questions. Int. J. Appl. Basic Med. Res. 10 (3), 210. https://doi.org/10.4103/ijabmr.IJABMR_30_20.

Gupta, V., Williams, E.R., Wadhwa, R., 2021. Multiple-choice tests: A–Z in best writing practices. Psychiatr. Clin. N. Am. 44 (2), 249–261. https://doi.org/10.1016/j.psc.2021.03.008.

Hamamoto Filho, P.T., Bicudo, A.M., 2020. Improvement of faculty's skills on the creation of items for progress testing through feedback to item writers: a successful experience. Rev. Bras. Educ. Méd. 44 (1). https://doi.org/10.1590/1981-5271v44.1-20190130.ing.

Hamamoto Filho, P.T., Silva, E., Ribeiro, Z.M.T., Hafner, M. de L.M.B., Cecilio-Fernandes, D., Bicudo, A.M., 2020. Relationships between Bloom's taxonomy, judges' estimation of item difficulty and psychometric properties of items from a progress test: a prospective observational study. Sao Paulo Med. J. 138 (1), 33–39. https://doi.org/10.1590/1516-3180.2019.0459.r1.19112019.

Hamamoto Filho, P.T., Bicudo, A.M., Pereira-Júnior, G.A., 2023. Assessment of medical students' surgery knowledge based on progress test. Rev. Col. Bras. Cir. 50. https://doi.org/10.1590/0100-6991e-20233636-en.

Karthikeyan, S., O'Connor, E., Hu, W., 2020. Motivations of assessment item writers in medical programs: a qualitative study. BMC Med. Educ. 20 (1), 334. https://doi.org/10.1186/s12909-020-02229-8.

Kelley, T.L., 1939. The selection of upper and lower groups for the validation of test items. J. Educ. Psychol. 30 (1), 17–24. https://doi.org/10.1037/h0057123.

Kirnbauer, B., Avian, A., Jakse, N., Rugani, P., Ithaler, D., Egger, R., 2018. First reported implementation of a German-language progress test in an undergraduate dental curriculum: a prospective study. Eur. J. Dent. Educ. 22 (4), e698–e705. https://doi.org/10.1111/eje.12381.

Kowash, M., Alhobeira, H., Hussein, I., Al Halabi, M., Khan, S., 2020. Knowledge of dental faculty in gulf cooperation council states of multiple-choice questions' item writing flaws. Med. Educ. Online 25 (1). https://doi.org/10.1080/10872981.2020.1812224.

Lindgren, S., Argullos, J.L.P., Millan, J.R., 2024. Assessment of clinical competence of medical students: future perspectives for Spanish Faculties. Med. Clín. Práct. 7 (2), 100424. https://doi.org/10.1016/j.mcpsp.2023.100424.

Malta, M., Cardoso, L.O., Bastos, F.I., Magnanini, M.M.F., Silva, C.M.F.P. da, 2010. Iniciativa STROBE: subsídios para a comunicação de estudos observacionais. Rev. Saúde Pública 44 (3), 559–565. https://doi.org/10.1590/S0034-89102010000300021.

McGahee, T.W., Ball, J., 2009. How to read and really use an item analysis. Nurse Educ. 34 (4), 166–171. https://doi.org/10.1097/NNE.0b013e3181aaba94.

McLaughlin, K., Lemaire, J., Coderre, S., 2005. Creating a reliable and valid blueprint for the internal medicine clerkship evaluation. Med. Teach. 27 (6), 544–547. https://doi.org/10.1080/01421590500136113.

Miller, G.E., 1990. The assessment of clinical skills/competence/performance. Acad. Med. 65 (9), S63–7. https://doi.org/10.1097/00001888-199009000-00045.

Neeley, S.M., Ulman, C.A., Sydelko, B.S., Borges, N.J., 2016. The Value of progress testing in undergraduate medical education: a systematic review of the literature. Med. Sci. Educ. 26 (4), 617–622. https://doi.org/10.1007/s40670-016-0313-0.

Sakai, M.H., Ferreira Filho, O.F., Matsuo, T., 2011. Avaliação do crescimento cognitivo do estudante de Medicina: aplicação do teste de equalização no Teste de Progresso. Rev. Bras. Educ. Méd. 35 (4), 493–501. https://doi.org/10.1590/S0100-55022011000400008.

Shaikh, S., Kannan, S.K., Naqvi, Z.A., Pasha, Z., Ahamad, M., 2020. The role of faculty development in improving the quality of multiple-choice questions in dental education. J. Dent. Educ. 84 (3), 316–322. https://doi.org/10.21815/JDE.019.189.

Tio, R.A., Schutte, B., Meiboom, A.A., Greidanus, J., Dubois, E.A., Bremers, A.J.A., 2016. The progress test of medicine: the Dutch experience. Perspect. Med. Educ. 5 (1), 51–55. https://doi.org/10.1007/s40037-015-0237-1.

Troncon, L.E. de A., Elias, L.L.K., Osako, M.K., Romão, E.A., Bollela, V.R., Moriguti, J.C., 2023. Reflections on the use of the Progress Test in the programmatic student assessment. Rev. Bras. Educ. Méd. 47 (2), e076. https://doi.org/10.1590/1981-5271v47.2-2022-0334.ing.

Villela, E.F. de M., Hyppolito, M.A., Moriguti, J.C., Bollela, V.R., 2022. Análise da adequação dos itens do Teste de Progresso em medicina. Rev. Bras. Educ. Méd. 46 (1). https://doi.org/10.1590/1981-5271v46.supl.1-20220303.

World Health Organization, 2021. Global Strategic Directions for Nursing and Midwifery 2021–2025. WHO. ⟨https://apps.who.int/iris/handle/10665/344562⟩.