

How assessment analytics can help to improve reliability, efficiency, and fairness of entrance examinations

Mika Nissinen¹, Elisa Silvennoinen¹ and Mohammed Saqr²

¹ Faculty of Social Sciences and Business Studies, Law School, University of Eastern Finland, Joensuu, Finland

² Faculty of Science and Forestry, School of Computing, University of Eastern Finland, Joensuu, Finland

Abstract

The study examined the level of difficulty, discrimination, and reliability of multiple-choice questions [MCQs] in Finnish national Law entrance examinations. The purpose was to assess whether MCQs could be used to rank applicants in a sufficiently reliable manner. The data set consists of anonymized scores from three exams (years 2018, 2019, 2021) containing 11,201 applicants all together. The study found that the MCQs provide a reliable, adequate, and high-quality discrimination. MCQ scores were also shown to correlate with essays and total scores.

Keywords

Assessment analytics, entrance examine, multiple choice question

1. Introduction

1.1 Background

Multiple-choice questions (MCQs) are a standard assessment modality across the full gamut of education disciplines. Their efficiency and the reliability have helped make MCQs almost ubiquitous in high and low stakes exams as well as formative and summative assessments. Besides, MCQs have an overall good correlation with the examined competences with a robust evidence that they measure the performance they are assumed to measure [1, 2, 3, 4]. MCQs are easy to administer, grading can be automated with barely any mistakes with the modern machines and on computers. Such features enable educators to assess a large number of students with the advantage of fast accurate grading with minimal human intervention [1, 5]. The benefits of MCQs include objective and unbiased grading, near-immediate results due to automation, and low cost. Handwriting speed does not affect the pace of answering the questions. MCQs enable a larger number of questions that can potentially better cover the study material. MCQs are particularly well suited for evaluating learners' starting levels [see for example 6, 7, 8, 9, 10, 11].

The benefits of essays may have been cited as potentially better evaluation of higher-order cognitive skills, more realistic account of real-life problems where different answers may be "correct" from different perspectives, and less room for guesswork [see for example 12, 13, 14, 15, 16]. A possible approach to compare essays and MCQs is to compare the results of testing the same content within the same context. Some studies have suggested that learners adapt their study methods by the type of questions they expect, and thus the cognitive depth of learning. Yet, other studies have suggested that MCQs facilitate more successful learning strategies [17].

Despite the many advantages MCQs offer to educators, reliable MCQs are challenging to construct as well as take a considerable time to author and review [1, 4, 18, 19]. The optimum MCQ requires item authors to follow certain guidelines, have hands-on training and fine tune their skills [18]. The authored MCQs are typically reviewed for language, scientific content and alignment with course objective [2, 20]. Following the examination, a review of item performance is often performed to assess the psychometric properties of questions and assess which ones fared below the expected level

Proceedings of the Finnish Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC22), Sept 29–30, 2022, Joensuu, Finland

EMAIL: mika.nissinen@uef.fi (A. 1); elisa.silvennoinen@uef.fi (A. 2); mohammed.saqr@uef.fi (A. 3)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

of quality [3, 4, 21, 22]. Therefore, the reliability of MCQs is reliant on good practices, and exam construction [2].

Learning analytics has emerged over a decade ago to use digital data to understand learning, learners and the environment in which learning occurs. Even though learning analytics have essentially covered all aspects of education, research on assessment continues to lag behind [23, 24]. One reason may be the perception that assessment is always regarded as an outcome that we must optimize: we put much effort to improve students' grades, rather than improve the way we score students' exams. In this study, we use analytical techniques along with the common psychometric methods to research assessment.

1.2 Motivation of the study

Entrance examination should rank the applicants reliably, fairly, and efficiently. Organizing the exam should not be exhaustive to universities and teachers, neither should taking the exam burden the applicants. This study is aimed at investigating the benefits and drawbacks of using MCQs in Finnish Law entrance examinations.

The study uses the assessments of three years of Finnish law schools' entrance exam, and apply analytics methods to examine the reliability of the exams, compare different exam modalities and try to answer whether MCQs are reliable, and most importantly, whether essays are needed [23, 24]. The results will be useful for evaluating the aims, quality, and fairness of entrance examinations as well as its challenges and how it can be improved. If the functionality of the multiple-choice questions in the entrance exams can be demonstrated, the multiple-choice questions can also be utilized effectively in the university education.

This study aims to analyze the psychometric properties of Finnish exams regarding difficulty, discrimination, and reliability. The overarching research question is how good and reliable are the MCQs and to what extent can they alone determine the student acceptance to law school? The following sub-questions were investigated:

- 1) how well do MCQs perform in assessing applicants?
- 2) how reliable are the MCQ -questions?
- 3) how does an applicant's MCQ score predict the overall score in the exam?

2. Materials and Methods

2.1 Context

In 2018 Finnish universities introduced multiple choice questions (MCQs) systematically in Law entrance examinations. Before that, the examinations used mostly essay questions. Entrance examination is the primary mode of entry to Law studies: 60% of entrants are drawn from entrance examinations, which consist of two parts: MCQs and essays. The grading is norm-referenced: Each applicant is compared to each other applicant, with the aim to rank-order the applicants. Grading is divided to two stages: First, applicants are ranked by their MCQ scores, automatically checked. The number of essays graded manually in the second stage is roughly thrice the annual intake.

As one of the most popular subjects of study, Law entrance examinations attract roughly 4000 applicants each year, with low acceptance rates [25]. For example, in 2021, 341 applicants were selected by the entrance exam, which is about 10 percent of the participants in the entrance exam [26]. Preparing for the examinations is a massive undertaking for the applicants, as well as the Finnish faculties of Law, who prepare study material, examination questions, and grading rubrics. Each year the student selection leads to hundreds of appeals, first to faculties of Law and some of them to administrative courts, which burdens the universities involved.

2.2 Materials

The data set consists of anonymized scores from the Finnish national Law entrance examinations from the years 2018, 2019, and 2021. Results from the year 2020 were excluded due to an

incommensurable exam set-up following the COVID-19 situation that year. The study included data for 11,201 applicants (2018 n= 4090, 2019 n= 3876, 2021 n=3235).

The study does not analyze the difficulty or discrimination of essay questions, because the essays are graded only for the top ≈40% of applicants. In addition, essay scores could not be easily compared with MCQ scores, because the number of essay questions is just four.

2.3 Methods

The analysis was carried out using item analysis as well as analytics of exams and students' performance. Item analysis is a well-established toolset for evaluation of the MCQs and exams in general. The analysis aimed at evaluation of the individual items, their difficulty, discriminating power of upper and lower achievers and the reliability of the exam. The main purpose of the entrance exam is to rank the candidates i.e., to distinguish between the good and the weak results effectively [See for example 7, 27, 28]. The Difficulty level, the discrimination index (DI), and the RIT index (item total correlation) were calculated for both, the individual MCQs and the total MCQ. In addition, an alpha drop value was calculated for each individual question. Alpha drop value for an individual MCQ indicates the residual reliability of all the MCQs if this individual MCQ was removed. Deleting a single question would not affect the reliability of the complete MCQ exam.

Difficulty level refers to the proportion of applicants who answered the question correctly. The value for difficulty level ranges from 0 to 1. The lower the level, the more difficult an individual question has been. A commonly used classification for difficulty level is: ≤0.50: hard; 0.51-0.84: moderate; >0.85: easy. Moderate questions are most appropriate. Too hard question may discriminate better between the high achievers (score distribution is skewed to the left) and too easy question between the low achievers (score distribution is skewed to the right) [29, 30].

Discrimination index indicates how well the question discriminates between high and low achieving applicants. The method used was the upper-lower index, which calculates the difference of the ratio of correct answers between the top and lower thirds (33%) of student exam. An item is considered discriminatory if its Discrimination index (DI) for the correct answer is positive (the value can be between +1 and -1). A negative value indicates a non-functioning item. The higher the value, the better the question discriminates. If the discrimination index is very low (or negative), the question cannot be considered discriminatory, as both successful and less successful candidates in the exam have answered the question.

The **RIT** (item total correlation) index indicates the correlation between an individual question and the entire entrance exam. The correlation ranges between +1 and -1; the higher the positive correlation, the better. This is directly related to the reliability. When a greater number of questions have a high RIT index, then also the reliability for the test is higher. The distinctiveness can be assessed on the basis of the RIT index [30]: <0.10: bad question that may need to be improved or removed; 0.10-0.19: weak question that should be improved; 0.20-0.29: acceptable, which can be improved; ≥0.30: Good.

The reliability estimate was also calculated for the entrance exam. Reliability refers to the consistency of measurements. Measurement results are considered reliable if they are stable and consistent between measurements. Reliable results are accurate, reproducible, and generalizable to other tests and similar metrics. The reliability means that the exam under similar conditions but in a different situation (= second entrance exam) leads to substantially same result. The reliability estimate was calculated using the split-half method, i.e., the questions were divided into two halves and the Cronbach's alpha factor was calculated for each. The process was repeated 10,000 times and results were averaged. The following classification can be used for reliability: ≤0.50: questionable reliability, test to be checked; 0.50-0.60: test revision recommendation; 0.60-0.70: somewhat low; 0.70–0.80 good; 0.80-0.90: Very good; <0.90: Excellent reliability [30, reliability, and Cronbach's alpha, see also for example 31, 32, 33, 34].

3. Results

For multiple choice questions (MCQ), the reliability estimate, difficulty level, and discrimination index were calculated. For the essay questions, these were not calculated, as only four of them per year is too few for reliable measurement results. In 2018 the mean reliability estimate for MCQs ($n = 40$) was 0.91 and in 2019 it was 0.94, while in 2021, the reliability was 0.88. The limit for excellent reliability is 90, which was exceeded in 2018 and 2019. The slightly lower reliability in 2021 maybe due to the fact that the MCQ exam questions were only 30. The mean item difficulty (P) values indicate moderate MCQs. The difficulty level in each exam (0.74 in 2018; 0.73 in 2019; 0.56 in 2021), is at the upper limit of moderate difficulty, which can be considered good. The higher the difficulty level, the easier the questions. The mean item discrimination index per each exam is positive, and the values show that the MCQs were discriminating (2018 0.34; 2019 0.43; 2021 0.42). A summary of means for reliability, difficulty and discrimination can be found in Table 1.

Table 1
Means for reliability, difficulty and discrimination

	Reliability*	Difficulty level	Discrimination (DI)
2018 MCQs (n=4090)	.91	.74	.34
2019 MCQs (n=3876)	.94	.73	.43
2021 MCQs (n=3235)	.88	.56	.42

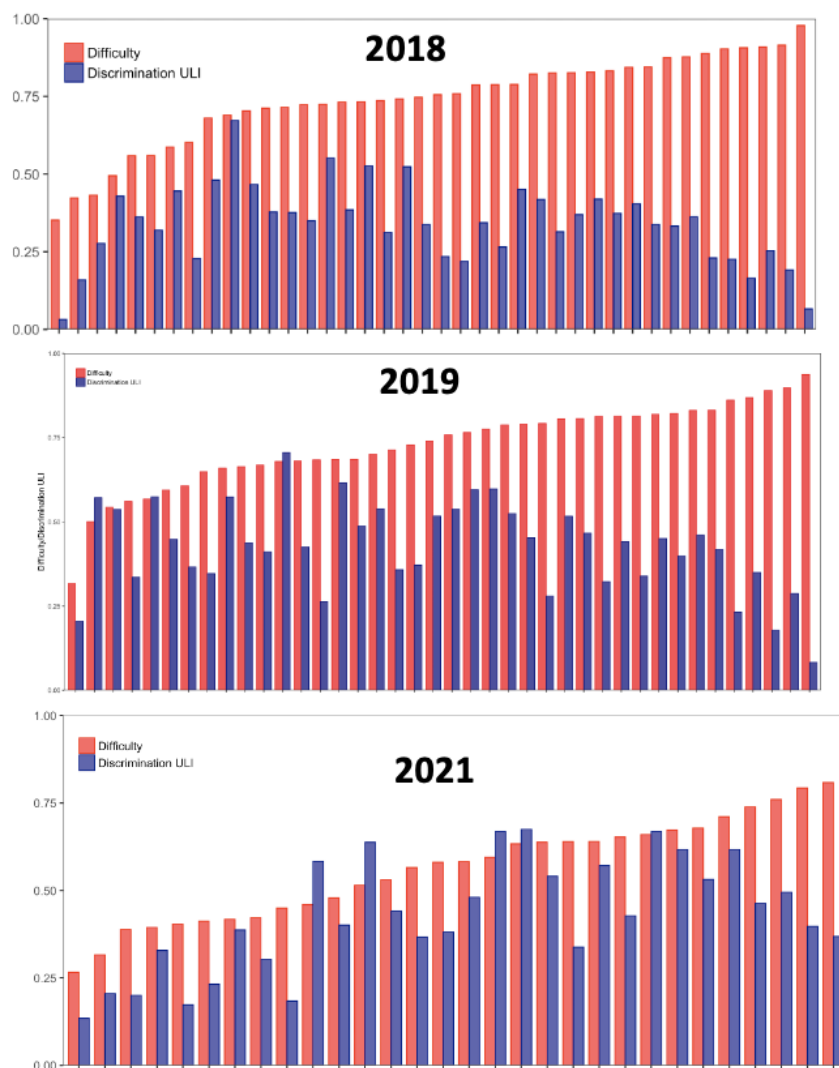


Figure 1: Item difficulty and discrimination of MCQs in 2018, 2019 and 2021 entrance exams.

Figure 1 shows the item difficulty and discrimination values of all individual MCQs in the 2018 (n=40), 2019 (n=40) and 2021 (n=30) entrance exams. In Fig. 1 red bars indicate the difficulty values and blue bars the discrimination values. The questions are ranked by the difficulty value. The first is the most difficult question and the last is the easiest question.

Figure 1 shows that MCQs have been balanced in terms of difficulty and discrimination. The 2021 exam contained a greater number of difficult items than the 2018 and 2019 exams did. However, there were not any very difficult ($P \leq 0.20$) items in any of the three exams. The 2018 and 2019 exams contained some easy ($P > 0.85$) items which almost all applicants were able to answer correctly.

The DI describes the difference between the best one-third and the weakest one-third. The value obtained from all questions were more than 0, and therefore, none of the questions were dysfunctional.

RIT index describes the discrimination of a single question in more detail than DI does. RIT index for nearly all the MCQs (n=110) is greater or equal than 0.30. This means either good or very good resolution. Only three MCQs has a RIT index less than 0.30, which is excellent result. A high index number indicates a strong positive correlation between the question and the total score in MCQs.

The reliability coefficient (Cronbach's alpha) for 2018 MCQs was 0.91, CI [0.9099–0.9101] and for 2019 MCQs was 0.94, CI [0.9399–0.9401]. These mean excellent reliability: that is, the MCQs in all exams measure the same thing and the candidates who answered a question correctly are likely to have answered the other questions correctly. In 2021, the reliability was slightly lower 0.88, CI [0.87–0.88].

Alpha drop for individual questions was between 0.90 and 0.91 (2018), 0.93 (2019) and between 0.87 and 0.88 (2021).

In addition to the psychometric properties the analysis shows how well does an applicant's MCQ score predict the overall score (MCQs and essays) in the exam. In this analysis only the number of admitted applicants and the minimum accepted scores were considered. The minimum accepted scores for admittance were 57 (2018), 55 (2019) and 42 (2021). The applicants were divided into accepted ("In") and rejected ("Out"). The original division between In and Out was based on the overall scores (MCQs and essays). However, for this analysis the Ins and the Outs were determined only by the MCQ scores. The main reason was to test how possible that an applicant gets selected based only on MCQ results. Figures 2 shows a multiway contingency plot, in which blue color represent a positive association and the size of the tile represent the magnitude. The chart shows the association between those who scored high in MCQs (MCQIn) and the entire entrance exam (ExamIn). Blue color describes applicants whose admittance status would not have changed had the selection been made by the total MCQ scores or the overall exam score (MCQs and essay). The top blue section shows applicants who scored high on both MCQs and the entire entrance exam and would thus be selected on the basis of both scores (MCQIn / ExamIn). The top red section describes applicants who received low scores on MCQs but high scores on the entire exam, i.e., performed better on essay questions. The lowest red section describes applicants who were successful in MCQs but less so in essay questions and were not selected on overall scores. The large lowest blue proportion describes applicants who received low scores from both MCQs and the complete exam and were not admitted.

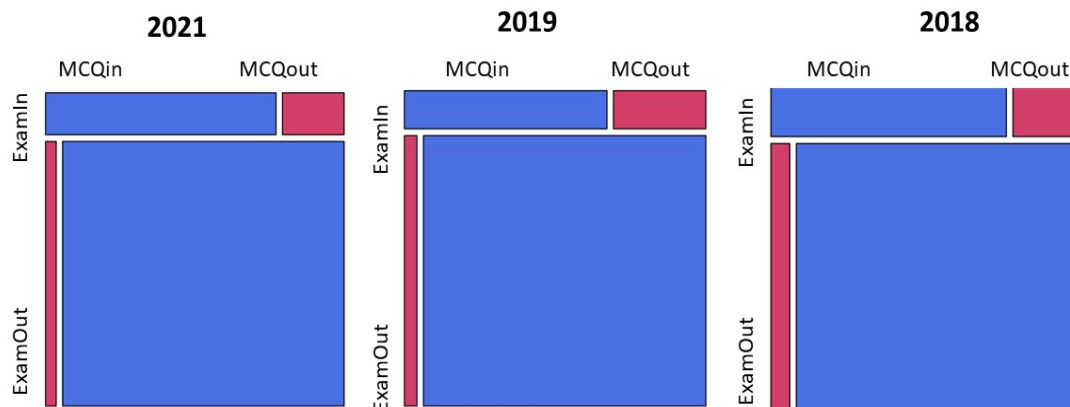


Figure 2: Agreement between two modes of student election: MCQs and overall score.

Figure 2 shows that applicants with high scores on MCQs are also highly likely to succeed in the entire entrance exam and therefore, high scores in MCQs predicts very well the overall score and eventually the admission. The analysis also calculated the odds ratio for those selected on the basis of MCQs and among the top in the complete entrance exam, which indicates association between the two factors. The odds ratio calculates the probability of being admitted divided by the probability of not being admitted. The ratio includes the probability of both admission and rejection. In 2019, the odds ratio was 49, in 2018 it was 53, and in 2021 the odds ratio was the highest 98.48.

4. Discussion

The results show that in 2018, 2019 and 2021 MCQs were able to divide high and low achievers reliably and consistently over the years. The analysis shows that MCQs are sufficient for student selection, and essays did not bring additional value that would significantly change the results, and therefore, the results do not support suggestions that MCQs and essays would measure very different sets of competences—or at least that there would be a significant number of students who would excel in one but not the other. Taken together, it seems that MCQs are far more reliable, consistent and offer a fair opportunity for every student to be graded the same way, free from human subjectivity (since essays are manually graded) as well as have extremely low potential for making grading mistakes. Therefore, we conclude that in order to offer applicants a fair equal opportunity, MCQs seems to be the exam modality of choice.

Taken together, the results of this study offers a proof of concept of how analytics results can possibly help offer a fair opportunity to applicants. Proving that objective types of examinations are reliable, reproducible and more differentiating paves the way for future adoption of such more reliable methods of assessments.

The results suggest that there is room for increasing the weight given to MCQs in the Finnish national Law entrance examinations. In addition to the psychometric results that support their use, increasing their share is supported by four arguments.

Firstly, the entrance examination is not aimed at evaluating expertise, but it is aimed at evaluating the potential and capacity of applicants for completing a LLB/LLM (Bachelor / Master of Laws) degree. The exam is not used to grade applicants but to divide them into two groups (accept/reject) using fair and objective criteria. The exam is expected to evaluate, in a very restricted time slot, a wide range of the applicants' abilities, which purpose is served by a well-designed, large set of MCQs.

Secondly, MCQ grading can be automated, which speeds up the process and release of results. The fewer essays there are to grade, the less faculties of Law have to reserve time to labor-intensive manual assessment. Should the entrance examination be administered fully online, automation could be used even more effectively. An online MCQ exam avoids several issues with paper-based MCQs, such as unclear markings, erased and re-written tick marks, scanning paper sheets, and errors in manual transfer of data.

Thirdly, MCQ scoring is objective. Applicant selects the correct answer from a set of pre-set answers. Scoring essays, on the contrary, requires interpretation of the applicant's intention compared to a pre-set sample answer. Several graders evaluate more than a thousand heterogeneous essays, which necessarily introduces some subjective elements in the process.

Fourthly, less ambiguity, more transparent, and objective MCQs ensure equal treatment of applicants and the rights of the individual applicant. This can lead to much fewer appeals [see also 35]. Table 2 shows appeals in 2018, 2019 and 2021. Rows present the years, and columns present exam questions (MCQs are grouped in groups of ten, each representing one topic in the study material). The last column (others) refers to appeals that were not about any specific question.

Table 2

Appeals in Finnish national Law entrance examinations, by year and question number

Year	MCQ1	MCQ2	MCQ3	MCQ4	Essay1	Essay2	Essay3	Essay4	Other
2018	5	4	32	12	105	69	85	64	4
2019	13	7	16	17	89	94	33	28	13
2021	2	21	59	5	66	69	68	46	16

As can be seen from the table, the very large number of appeals related to essay questions, as compared to MCQs. This indicates that applicants experienced especially essays to be incorrect or unfair. Appeals generate massive amount of extra effort for applicants, faculties of Law, and possible even courts (if the appeal continues there). Due to the heterogeneity of essay answers, appeals related to essays are more time consuming than appeals related to MCQs. Increasing the share of MCQs in the exam would benefit all the parties due to improved trust in the objectivity and fairness of the process, and due to decreased number of appeals.

5. Conclusions

The analysis of more than ten thousand applicants has emphasized that the MCQs offer an objective, reliable and differentiating type of assessment. Adopting MCQs for entry exams has the potential for offering a fair opportunity for future applicants with less subjectivity.

Acknowledgements

The authors wish to thank Professor Matti Tedre for commenting the manuscript. The authors also wish to thank exam coordinator Kirsi Kaljunen for helping with the data collection.

References

- [1] Koediger, H. L., Marsh, E. J. (2005), The positive and negative consequences of multiple-choice testing, *Journal of Experimental Psychology, Learning, Memory, and Cognition*, 31(5), 1155–1159.
- [2] McCoubrie, P. (2004), Improving the fairness of multiple-choice questions: A literature review, *Medical Teacher*, 26(8), 709–712.
- [3] Tarrant, M., Ware, J. (2008), Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments, *Medical Education*, 42(2), 198–206.
- [4] Ware, J., Vik, T. (2009), Quality assurance of item writing: During the introduction of multiple choice questions in medicine for high stakes examinations, *Medical Teacher*, 31(3), 238–243.
- [5] Bailey, P. H., Mossey, S., Moroso, S., Cloutier, J. D., Love, A. (2012), Implications of multiple-choice testing in nursing education. *Nurse Education Today*, 32(6), e40–e44.
- [6] Photopoulos, P., Tsonos, C., Stavrakas I., Triantis, D. (2021), Preference for Multiple Choice and Constructed Response Exams for Engineering Students with and without Learning Difficulties, In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDEU 2021)*, Vol. 1, 220–231.

- [7] AlKhatib, H. S., Brazeau, G., Akour, A., Almuhaissen, S. A. (2020), Evaluation of the effect of items' format and type on psychometric properties of sixth year pharmacy students clinical clerkship assessment items. *BMC Medical Education* Vol. 20 Article 190, 1–8.
- [8] Kaipa, R. M. (2020), Multiple choice questions and essay questions in curriculum, *Journal of Applied Research in Higher Education*, Vol. 13 No. 1, 16–32.
- [9] Ul Islam, Z., Usmani, A. (2017), Psychometric analysis of Anatomy MCQs in Modular examination, *Pak J Med Sci*, Vol. 33 No. 5, 1138–1143.
- [10] Ali, S. H., Carr, P. A., Ruit, K. G. (2016), Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter, *Journal of the Scholarship of Teaching and Learning*, Vol. 16 No. 1, 1–14.
- [11] Kastner, M., Stangl, B. (2011), Multiple Choice and Constructed Response Tests: Do Test Format and Scoring Matter? *Procedia Social and Behavioral Sciences* Vol. 12, 263–273.
- [12] OECD (2013), *OECD Reviews of Evaluation and Assessment in Education: Synergies for Better Learning – An international perspective on evaluation and assessment*.
- [13] Buckles, S., Siegfried, J. J. (2006), Using Multiple-Choice Questions to Evaluate In-Depth Learning of Economics, *Journal of Economic Education*, Vol. 37 No. 1, 48–57.
- [14] Tozoglu, D., Tozoglu, M. D., Gurses, A., Dogar, C. (2004), The students' perceptions: essay versus multiple-choice type exams, *Journal of Baltic Science Education*, Vol. 3 No. 2, 52–59.
- [15] Kuechler, W. L., Simkin, M. G. (2010), Why Is Performance on Multiple-Choice Tests and Constructed-Response Tests Not More Closely Related?, *Theory and an Empirical Test*, *Decision Sciences Journal of Innovative Education*, Vol. 8 No. 1, 55–73.
- [16] Scouller, K. (1998), The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay, *The International Journal of Higher Education Research* Vol. 35 No. 4, 453–472.
- [17] Hift, R. J. (2014), Should essays and other “open-ended”-type questions retain a place in written summative assessment in clinical medicine? *BMC Medical Education*, Vol. 14 Article 249, 1–18.
- [18] Haladyna, T. M., Downing, S. M., Rodriguez, M. C. (2002), A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment, *Applied Measurement in Education*, 15(3), 309–333.
- [19] Schuwirth, L. W. T. (2003), ABC of learning and teaching in medicine: Written assessment, *BMJ*, 326(7390), 643–645.
- [20] DiBattista, D., Kurzawa, L. (2011), Examination of the Quality of Multiple-choice Items on Classroom Tests. *Canadian Journal for the Scholarship of Teaching and Learning*, 2(2), 1–23.
- [21] Mukherjee, P., Lahiri, S. K. (2015), Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. *IOSR Journal of Dental and Medical Sciences*, 14(12), 2279–2861.
- [22] Redmond, S. P., Hartigan-Rogers, J. A., Cobbett, S. (2012), High Time for a Change: Psychometric Analysis of Multiple-Choice Questions in Nursing, *International Journal of Nursing Education Scholarship*, 9(1), 1–16.
- [23] The authors have written a similar kind of analysis in Finnish for the statistics of 2018 and 2019, see Nissinen, M., Silvennoinen, E., Saqr, M. (2022), *Monivalintakysymykset oikeustieteellisen alan yhteisvalintakokeessa – hitti vai huti?*, *Edilex-article* 13.1.2022, 1–19.
- [24] Saqr, M, *Assessment analytics (2017): The missing step*, *Int J Health Sci (Qassim)*. Jan-Mar; 11(1), 1–2.
- [25] <https://oikeustieteet.fi/kokeet-ja-tilastot>.
- [26] <https://oikeustieteet.fi/valintakoevalinnan-tulokset-julkaistu-6-7-2021>.
- [27] Ul Islam, Z., Usmani, A. (2017), Psychometric analysis of Anatomy MCQs in Modular examination, *Pak J Med Sci*, Vol. 33 No. 5, 1138–1143.
- [28] Siri, A., Freddano, M., (2011), The use of item analysis for the improvement of objective examinations, *Procedia - Social and Behavioral Sciences* Vol. 29, 188–197.
- [29] https://www.washington.edu/assessment/scanningscoring__trashed/scoring/reports/item-analysis/.
- [30] Osterlind, S. J. (1998), *Constructing Test Items: Multiple-Choice, Constructed-Response, Performance, and Other Formats*, Kluwer Academic Publishers, Boston, 1998.

- [31] Tavakol, M., Dennick, R., (2011), Post-examination analysis of objective tests, *Medical Teacher*, Vol. No. 6, 447-458.
- [32] Bacon, D. R. (2003), Assessing Learning Outcomes: A Comparison of Multiple-Choice and Short-Answer Questions in a Marketing Context. *Journal of Marketing Education*, Vol. 25 No. 1, 31-36.
- [33] Cho, E. (2016), Making Reliability Reliable: A Systematic Approach to Reliability Coefficients, *Organizational Research Methods*, Vol. 19 No. 4, 651-682.
- [34] Cronbach L. J. (1951), Coefficient alpha and the internal structure of tests, *Psychometrika*, Vol. 16 No. 3, 297-334.
- [35] Tozoglu, D., Tozoglu, M. D., Gurses, A., Dogar, C. (2004), The students' perceptions: essay versus multiple-choice type exams, *Journal of Baltic Science Education*, Vol. 3 No. 2, 52-59.