Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

DETERMINING THE PERFORMANCE LEVELS OF EIGHTH GRADE STUDENTS IN THE NATIONAL TEST FOR CONTROLLING THE QUALITY OF EDUCATION IN SCIENCE USING LATENT CLASS ANALYSIS

GHADA FADEL SALAMEH SHATNAWI
Educational Sciences, Yarmouk University. Email: gfadel4@yahoo.com

Abstract

The purpose of the study was to determine the number of latent classes in the performance of eighth-grade students on the National Test for the quality of education in the fields of science, then to reveal the probability of answering each item of the science test by students across the different latent class levels, and then to study the extent of agreement between Latent Class levels and performance levels approved by the Ministry of Education. To achieve the aim of the study, the researcher analyzed the response patterns of eighth grade students on the and science test using the Mplus software. The study sample, on which statistical processing was applied, consisted of 3000 male and female eighth grade students in Jordan. The results of the study showed that there are three latent classes for the national science test based on students' response patterns. The probability of students answering the test items played an important role in differentiating between students' abilities in the form of latent classes and supported the three latent levels model. Finally, using the Chi-square test and contingecy coefficient, the study concluded that there is agreement between the latent performance levels and the performance levels approved by the Ministry of Education. The study concluded with a set of recommendations that there is a clear weakness in performance in the science test, and the reasons for such weakness are unknown, which necessitates the need to investigate whether this weakness reflects an actual weakness in the subject or not, and to reconsider the cut-off scores adopted by the Ministry of Education and adopt one of the known methods in determining cut-off scores instead of relying on the arbitrary method. The four levels used by at the Ministry of Education are not distinct in classifying students, as the basic and partial mastery levels can be combined together.

Keywords: Performance Standards, Performance Levels, Cut Score, Latent Trait, Response Pattern, Latent Class Analysis.

INTRODUCTION

The social and psychological sciences have recently gained increased interest in latent class analysis, as these sciences rely on numerous population groups in which individuals within each group are largely similar in characteristics, while individuals across different subgroups vary. Statistical inferences assuming homogeneity across population groups are misleading, known as Simpson's paradox (Agresti, 1996; Simpson, 1951). If the source of heterogeneity within the groups is known, such as gender, and this is confirmed by statistically comparing subgroups by gender, then the statistical methods that can be applied include: t-test, ANOVA, MANOVA, regression, and multigroup factor analysis. If the source of heterogeneity among subpopulations is unknown and cannot be distinguished based on observed characteristics, the subpopulations are called latent classes. Appropriate statistical methods for detecting these groups include latent class

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

analysis (LCA), cluster analysis, latent profile analysis, and finite mixture models (Muthén & Muthén, 1998). Both cluster analysis and LCA can be used to classify similar participants into groups or classes, but LCA has several advantages: First, the number of groups is random in cluster analysis, whereas the theoretical formulas for each class in LCA can be directly specified and empirically tested against the dataset. LCA allows for more rigorous methods for comparing alternative models, such as likelihood ratio tests, Akihiki information criterion (AIC), and Bayes information criterion (BIC). Second, latent class analysis (LCA) is a robust measure of the various observed variables, which is always a problem with cluster analysis. Third, LCA takes into account the uncertainty of an individual's membership in a latent class, whereas cluster analysis cannot (Vermunt, 2002 & Magidson). Thus, LCA is a measurement model that enables us to qualitatively classify individuals into classes, groups, or latent classes based on their response patterns to a set of questionnaire questions, test items, or any set of observed variables to detect latent variance in samples. It assumes that qualitative variation between classes represents all relationships among the data (Hagenaars & McCutcheon, 2002). LCA can be thought of as a means of grouping similar individuals together, in contrast to factor analysis, which is a means of grouping similar items together. It is an empirically derived, "person-centered" approach, in contrast to the traditional, "variable-centered" approach, which generally requires arbitrary cutoff scores for classification or discrimination between individuals (Nylund et al., 2007).

In Latent Class Analysis (LCA), each respondent is assigned to a single latent class based on their observed response pattern. Under the assumption of local independence within each latent class, the probability of a correct response to a given test item is independent of responses to other items once the latent class membership is accounted for. Given a set of dichotomous items administered to N examinees, where $x_i = 1$ denotes a correct response to item i and $x_i = 0$ represents an incorrect response ($i = 1, 2, ..., n$), let P_{iK} represent the probability of a correct response to item i within latent class k. Consequently, the probability of an incorrect response is $1 - P_{iK}$. If the data structure assumes K latent classes, the probability of observing response pattern r, denoted as P_r, is given by:

$$P_r = \sum_{K=1}^{K} \pi_K \prod_{i=1}^{n} P_{iK}^{x_i} (1 - P_{iK})^{1 - x_i}$$

where:

- $P(r \mid K)$ is the probability of response pattern r conditional on membership in latent class K.

- π_K represents the class membership probability, which indicates the proportion of individuals classified into latent class K.

When conducting LCA on empirical data, both $P(r \mid K)$ and π_K must be estimated (Hagenaars & McCutcheon, 2002). If these parameters are treated as free estimates, the latent class model is referred to as an unrestricted model, which may lead to issues of non-identifiability, meaning that a unique set of parameter estimates may not exist (Vermunt & Magidson, 2000).

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Traditional LCA seeks to extract the optimal number of latent classes needed to account for all item response dependencies. For example, if a test consists of four items (j = 4), the response probability can be expressed as:

$$P(y_1, y_2, y_3, y_4) = \sum_{K=1}^{K} P(X = k) P(y_1, y_2, y_3, y_4 \mid X = k)$$

where:

- $P(X = k)$ denotes the probability of an individual belonging to latent class k.

- $P(y_1, y_2, y_3, y_4 \mid X = k)$ represents the conditional probability of a specific response pattern given membership in class k.

This equation reflects the principle that each latent class is characterized by distinct item response probabilities, and the overall response distribution is modeled as a weighted mixture of the latent classes, with $P(X = k)$ serving as the mixture weights. A fundamental assumption of LCA is that the latent variable explains all observed associations among the test items, formalized through the assumption of local independence (Vermunt & Magidson, 2004). Mathematically, local independence is expressed as:

$$P(y_1, y_2, y_3, y_4 \mid X = k) = P(y_1 \mid X = k) P(y_2 \mid X = k) P(y_3 \mid X = k) P(y_4 \mid X = k)$$

By substituting the local independence assumption into the class-conditional response model, we obtain the standard LCA probability structure:

$$P(y_1, y_2, y_3, y_4) = \sum_{K=1}^{K} P(X = k) P(y_1 \mid X = k) P(y_2 \mid X = k) P(y_3 \mid X = k) P(y_4 \mid X = k)$$

The posterior probability of class membership given response pattern r is computed as:

$$P(k \mid r) = (\pi_K \prod_{i=1}^{n} P_{iK}^{x_i} (1 - P_{iK})^{1 - x_i}) / P(r)$$

for $1 \leq k \leq K$.

In most software applications for mixture modeling, LCA parameters are estimated using Maximum Likelihood (ML) or Expectation-Maximization (EM) algorithms. ML estimation is particularly useful in handling item-level missing data under the Missing at Random (MAR) assumption, as cases with partial missingness can still contribute to the likelihood function (Nylund-Gibson & Choi, 2018). A common issue in mixture models is that the likelihood function may converge to a local rather than a global maximum (McLachlan et al., 1999). To mitigate this, multiple sets of random starting values are used, and models are re-estimated to verify convergence to a consistent solution across different runs (Berlin et al., 2014; Masyn, 2013). This approach is automated in software such as Mplus, where users can specify the number of random starting values for optimization. The Bayesian estimation method is widely used in the social sciences (Kaplan & Vioante, 2014) due to its simplicity—especially when not all model assumptions (such as conditional independence) are met. Using prior information about the model, researchers can more accurately specify small correlations between items within the same class and approximate the dependencies within that class without needing to restructure the emerging classes (Asparouhov & Muthén, 2015). Studies have shown that when accurate

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

and rich prior information is available, its use reduces bias in parameter estimation and enhances class identification (Depaoli, 2012; Depaoli et al., 2017). Latent Class Analysis (LCA) for setting performance standards differs from other standard-setting methods in many fundamental assumptions. It does not assume the existence of a continuous trait to explain performance; rather, it relies on response patterns and the fit between data and estimated parameters across models with different numbers of latent classes. These models are then tested to ensure that the selected latent class model adequately represents the relationships in the given data (Dayton, 1991; Haertel, 1984, 1989; Luecht & DeChamplain, 1998).

The educational literature highlights many applications of LCA in identifying performance levels across various tests. For example, Brown (2007) applied several latent class models to student responses to 10 multiple-choice math items, and the analysis revealed two latent classes that could explain the variation in student performance. He recommended using LCA to analyze response patterns and judge student proficiency levels instead of relying on costly traditional methods that depend on expert judgment. Cogo-Moreira et al. (2013) found that the best latent class model to explain differences in reading and writing tests was a three-class model. Similarly, Jarar and Bani Ata (2018), in their analysis of Jordanian students' performance on the 2011 TIMSS mathematics test, found three latent classes influenced by the teacher's gender, the students' gender, and the type of school—favoring students from private schools. Sideridis et al. (2021) used multilevel latent class analysis (Schmiege et al., 2017; Mäkikangas et al., 2018) to identify performance levels of Saudi high school students based on their demographic characteristics, parental characteristics, and school-related behaviors such as absenteeism. The analysis revealed four latent classes, based on several criteria and fit indices (Masyn, 2013). Parental education and the number of student absences significantly influenced class membership, acting as positive and negative predictors of achievement levels in the latent classes.

The use of LCA has not been limited to testing. For example, Bani Ata (2022) applied latent class analysis to data from the Harrison and Bramson thinking styles scale, standardized for the Arab environment. The analysis revealed three latent classes: the analytical style (32%), the idealistic style (46%), and the realistic-practical style (22%). The results showed a relationship between the latent classes and students' academic degrees.

Study Problem and Questions:

The development of the educational system, training programs, new curricula, and modern teaching and assessment strategies all aim to improve the quality of education in the Hashemite Kingdom of Jordan. As part of this development process, the Ministry of Education, through the Department of Examinations and Testing, administers the National Test for Monitoring Education Quality and determining students' performance levels. There is a general agreement on classifying students into four performance levels: Basic, Partial Mastery, Full Mastery, and Advanced. Arbitrarily, three cut scores on a percentage scale—30, 50, and 70—are used to separate each pair of consecutive levels.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

However, it appears that the process of determining performance levels has not been based on any established standard-setting methods known in educational literature. Instead, it is a subjective process that depends solely on a percentage of the individual's total score, without considering the response patterns to individual test items. Therefore, the current study aims to determine performance levels based on response patterns using Latent Class Analysis (LCA). The goal is to identify the performance levels of eighth-grade students who took the National Science Test, using LCA. This will help in understanding the characteristics of different groups and students' response patterns, and in adopting a comprehensive educational approach that considers the needs and responses of each latent class. Furthermore, it encourages collaboration between teachers and educational supervisors to exchange expertise and effective teaching strategies based on students' classification into latent classes, and to analyze the impact of such classification on the development of tailored instructional programs to meet the needs of each group.

Accordingly, the study seeks to answer the following two questions:

1. How many latent performance levels exist among eighth-grade students on the National Test for Monitoring Education Quality in Science?
2. What are the probabilities of responding to each science test item by students across the different latent performance levels?

Significance of the Study:

The practical significance of the study lies in classifying students into performance levels based on their actual responses, which enables a clear description of what students at each level know and can do. This is beneficial in identifying the strengths and weaknesses associated with each level. Based on these findings, effective remedial plans can be developed to address areas of weakness and meet the students' most urgent needs. From a theoretical perspective, the importance of the study lies in the use of Latent Class Analysis (LCA) to determine student performance levels. This opens the door for educators to apply this method in various educational issues and contexts.

Operational Definitions and Concepts:

- Performance Standards: Performance standards are tools in the form of a scale used to interpret quantitative data qualitatively. They describe a student's performance in achieving a specific objective and serve as an effective tool for assessing the extent to which the results align with predetermined goals (Wilde & Pieter, 2018).
- Performance Levels: These are the categories into which students are classified based on the performance standards used after conducting tests and assessments. Each performance level is characterized by a certain amount of information regarding students' mastery of the required performance skills (García & Palomares, 2021).
- Cut Score: Hambleton (1978) defined cut scores as "dividing points on a continuous performance scale, where test results are divided into different categories." A cut score refers to the minimum level of competency required for an individual to be classified

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

into a certain level (e.g., good, average, weak) in criterion-referenced tests (Allam, 2007; Crocker & Algina, 1986).

- Response Pattern: The individual's responses to test items in a way that reflects their level of knowledge, attribute, or skill, which in turn determines their performance level on the test (Nylund-Gibson & Choi, 2018).
- Latent Class Analysis (LCA): A measurement model that enables the qualitative classification of individuals into groups or latent classes based on their response patterns to a set of questionnaire items, test items, or any observed variables. It helps uncover the underlying heterogeneity in the sample and assumes that the qualitative differences between the classes explain all the relationships within the data (Vermunt & Magidson, 2005).

Study Limitations:

- This study was limited to the data provided by the Examinations Department at the Ministry of Education regarding the National Test for Monitoring Education Quality in the subject of Science for the academic year 2021/2022, paper-based version only.
- The study was restricted to eighth-grade students only.

Previous Studies:

In 2007, Brown (2007) applied a student assessment tool consisting of 10 multiple-choice items and two performance tasks to a sample of 191 seventh and eighth-grade students in mathematics. The researcher analyzed students' results and classified them qualitatively using Latent Class Analysis (LCA). Multiple latent class models were used based on dichotomous items. The analysis revealed the presence of two latent classes that could explain the variation in student performance. The researcher also concluded that experimental methods such as LCA, which rely on response patterns, can be used to assess student proficiency instead of costly traditional judgment-based methods. Cogo-Moreira et al. (2013) aimed to identify the best latent class model using LCA on the reading and writing subtests of the Academic Performance Test (TDE). The researchers selected a sample of 1,945 students, aged between 6 and 14 years, with IQ scores above 70, from public schools in São Paulo (35 schools) and Porto Alegre (22 schools). The researchers analyzed the subtest results using LCA and identified three latent classes, which demonstrated good discrimination and explanatory power for the data. They also concluded that experimental methods such as LCA are effective in accurate classification.

Jarar and Bani Ata (2018) selected test booklet No. 11 from the 2011 TIMSS mathematics assessment to identify the number of latent classes that differentiate between Jordanian students' abilities based on the probability of answering number and algebra content items correctly. They also explored the demographic characteristics of students that contributed to latent class membership and investigated the underlying reasons behind Jordan’s decline in international ranking on the TIMSS mathematics test. The study sample consisted of 531 eighth-grade students from Jordan. The findings revealed three latent classes, where teacher gender, student gender, and school type played a role in

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

classifying students, favoring students from mixed private schools whose teachers reported that the curriculum did not adequately meet their cognitive needs. The study also found two latent classes for the algebra content domain, influenced by student gender, school location, and whether the teacher provided problem-solving explanations, benefiting urban female students who received such explanations.

Sideridis et al. (2021) selected three samples of 2,000 students each, drawn randomly from a large population of 500,000 students in Saudi Arabian schools across the years 2016, 2017, and 2018, aiming to identify high school students' academic achievement as a function of demographic characteristics, parental attributes, and school behaviors, such as absenteeism. The researchers used Multilevel Latent Class Analysis (MLCA) as developed by Schmiege et al. (2017) and Mäkikangas et al. (2018). The results showed the existence of four latent classes, based on information criteria such as BIC, Bayes, and other indices proposed by Masyn (2013). Parental education and student absenteeism significantly influenced classification as positive and negative predictors of achievement levels within the latent classes. Bani Ata (2022) used the Harrison and Bramson Thinking Styles Scale, standardized for the Arab context (1995), on a sample of 418 students from the Faculty of Education during the 2018–2019 summer semester. The results were analyzed using Latent Class Analysis, which revealed three latent classes:

- Analytical style (31.9%)
- Idealistic style (45.5%)
- Realistic-practical style (22.6%)

The findings indicated a relationship between the latent classes identified by LCA and students' academic degree levels.

Commentary on Previous Studies:

The review of previous studies indicates that **experimental methods**, such as **Latent Class Analysis (LCA)**, can be used to evaluate student proficiency instead of relying on costly traditional judgment-based methods, as demonstrated in the findings of **Brown (2007)** and **Cogo-Moreira et al. (2013)**.

In Bani Ata's (2022) study, the results showed a relationship between the latent classes identified through LCA and students' academic degree levels. Meanwhile, Vermunt and Magidson concluded that LCA significantly outperforms the K-means method, and that it is indistinguishable from Discriminant Analysis (DISC) in terms of classification capability (Magidson & Vermunt, 2002). Moreover, the reviewed studies highlighted LCA's effectiveness in classifying students into multiple performance levels based on their qualitative characteristics. For example, in the study by Sideridis et al. (2021), parental education and the number of student absences were found to significantly influence classification, acting as positive and negative predictors of achievement levels in the latent classes. Similarly, the results of the study by Jarar and Bani Ata (2018) showed that teacher gender, student gender, and school type played a role in classifying students

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

into three latent classes. In the algebra content domain, student gender, school location, and whether the teacher provided problem-solving explanations were found to be influential in classifying students into two latent classes.

Study Methodology:

The descriptive-analytical method was used to identify the number of latent classes in the performance of eighth-grade students on the National Test for Monitoring Education Quality in Mathematics in Jordan, as this method aligns with the nature and objectives of the research.

Population and Sample of the Study:

The study population consisted of all eighth-grade students in public, private, UNRWA (United Nations Relief and Works Agency) schools, and military education schools across the Hashemite Kingdom of Jordan during the 2021/2022 academic year, who actually sat for the National Test. The total number of students was 67,589.

The study sample consisted of 3,000 students, distributed as follows:

- 1,500 students who took the paper-based science test
- 1,500 students who took the electronic science test

The samples were randomly selected from the study population using Excel software, with students randomly chosen from all educational directorates affiliated with the Jordanian Ministry of Education. The selection was proportionally distributed based on each directorate's representation in the population, as shown in Table 1.

Table 1: Distribution of the Study Sample of Eighth-Grade Students Across Educational Directorates

Directorate	Number of Students Taking the Paper-Based Science Test	Number of Students Taking the Electronic Science Test
Private Education	105	102
Amman	400	409
Irbid	246	246
Jerash	40	40
Ajloun	39	33
Mafraq	99	99
Zarqa	201	210
Salt	89	89
Karak	62	62
Tafilah	21	21
Ma'an	24	24
Aqaba	32	32
UNRWA	121	121
Military Education	21	12
Total	**1500**	**1500**

4o

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Study Instrument:

To achieve the objectives of the study, the **Science Test** from the **National Test for Monitoring the Quality of Education** for **eighth-grade students** in the **Hashemite Kingdom of Jordan** for the academic year **2021/2022** was used.

The science test consists of **40 multiple-choice items** covering the following domains:

- Interrelationships within ecosystems
- Biodiversity
- Reproduction and genetics
- The human body and health
- Mechanics
- Electricity and magnetism
- Thermodynamics
- Waves
- Matter: structure and properties
- Earth and environmental sciences
- Astronomy and space science

The total score for the test is **40 points**. These items assess student performance in the following skills: **inquiry, classification, tracking processes, and life cycles** (Ministry of Education, 2022).

Table 2: Reliability Coefficients of the National Test for Monitoring Education Quality Using Cronbach's Alpha

Test Version	Science
Paper-Based	0.78
Electronic	0.79

The statistical analysis results revealed the means of student performance in science for eighth-grade students, along with standard deviations and the t-test value to detect the significance of gender-based differences in mean scores at the national level and across both test formats (paper-based and electronic), as shown in Table 3 (Ministry of Education, 2022).

Table 3: Mean Percent Scores of Eighth-Grade Students in Science, Corresponding Standard Deviations, and t-test Values According to Test Format

Test Format	Overall Mean Score (%)	Overall SD	Female Mean Score (%)	Female SD	Male Mean Score (%)	Male SD	t-value
Paper-Based	42	15	43	15	40	16	19.9
Electronic	46	16	47	15	44	16	7.0

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Based on the results of eighth-grade students' performance on the science test, they were **classified into performance levels** in the science subject **nationwide**, according to performance indicators, as shown in **Table 4**.

Table 4: Percentage Distribution of Eighth-Grade Students by Performance Levels in Science Nationwide According to Performance Indicators

Paper-Based	Electronic	Performance Level	Description
7%	9%	**Advanced Level**	Demonstrates mastery of all required knowledge and skills, and achieves learning outcomes exceeding the standards of the specified educational level.
21%	29%	**Full Mastery Level**	Demonstrates mastery of most required knowledge and skills, and achieves the learning outcomes for the specified educational level.
53%	49%	**Partial Mastery Level**	Demonstrates mastery of some required knowledge and skills and is approaching the achievement of learning outcomes for the specified educational level.
19%	13%	**Basic Level**	Does not demonstrate the minimum required knowledge and skills; requires a remedial plan to redirect learning on the right path.

Statistical Analysis:

To answer the study's two main questions regarding:

1. The number of latent performance levels among eighth-grade students on the National Test for Monitoring Education Quality in Science, and
2. The probabilities of responding to each science test item across the different latent performance levels,

Mplus software was used to conduct Latent Class Analysis (LCA) on students' responses. The number of students in each latent class was calculated, and the response probability for each of the 40 items was computed for each latent class.

Additionally, the following statistical indicators were calculated to assess the model fit quality of the latent class model to the data (Chen et al., 2017):

Bootstrapped Likelihood Ratio Test (BLRT), Voung-Lo-Mendell-Rubin (VLMR), and Lo-Mendell-Rubin Adjusted Likelihood Ratio Test (LMR) were used to assess model stability. Additionally, the following information criteria were calculated:

- Akaike Information Criterion (AIC)
- Bayesian Information Criterion (BIC)
- Adjusted Bayesian Information Criterion (Adjusted BIC)
- Entropy Index
- Scaling Correction Factor of MLR (Maximum Likelihood Robust Estimator)

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

These indicators were used to evaluate the goodness-of-fit and model quality in the latent class analysis.

RESULTS

First: What is the number of latent performance levels among eighth-grade students on the National Test for Monitoring Education Quality in Science?

The optimal number of latent classes for the paper-based and electronic versions of the National Science Test was determined using MPlus software and based on the parametric BLRT (Bootstrapped Likelihood Ratio Test) statistic, which is considered one of the best indicators for detecting the appropriate number of latent classes. Table 5 shows the BLRT values for each latent class model in both test formats.

Table 5: BLRT Parametric Statistic Values for Each Latent Class Model in the Paper-Based and Electronic Versions of the Science Test

Test	Format	No. of Latent Classes	No. of Parameters	2 * Log-Likelihood Difference	Difference in Parameters	Approx. Statistical Significance
Science	Paper-Based	3	122	883.567	41	0.0000
		4	163	232.071	41	0.0351
		5	204	124.207	41	0.5921
Science	Electronic	3	122	839.755	41	0.0000
		4	163	375.362	41	0.0338
		5	204	137.676	41	0.2743

As shown in Table 5, **two acceptable models** were identified for both the **paper-based and electronic** versions of the science test: the **three-class model** and the **four-class model**.

Additionally, the **VLMR (Vuong-Lo-Mendell-Rubin)** and **LMR (Lo-Mendell-Rubin adjusted LRT)** indicators were also used to determine the acceptable latent class model. Table 6 presents the values for the LMR and VLMR indicators.

Table 6: VLMR and LMR Values for Each Latent Class Model in the Paper-Based and Electronic Versions of the Science Test

Test	No. of Classes	No. of Parameters	VLMR Value	VLMR p-value	LMR Value	LMR p-value
Science (Paper-Based)	3	122	-36178.935	0.0000	88.630	0.0000
	4	163	-35737.151	0.0351	231.000	0.0359
	5	204	-35621.115	0.5921	123.794	0.5933
Science (Electronic)	3	122	-35970.055	0.0000	836.964	0.0000
	4	163	-35550.177	0.0338	374.115	0.0344
	5	204	-35362.496	0.2743	137.218	0.2756

As shown in Table 6, **two models** (the **three-class** and **four-class** models) are statistically acceptable for both the paper-based and electronic versions of the science test. Therefore, it was necessary to rely on additional **information criteria**, including **BIC,**

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

AIC, Adjusted BIC, Scaling Correction Factor for MLR, and Entropy, to determine the best-fitting model. These values are presented in **Table 7**.

Table 7: Information Criteria (BIC, AIC, Adjusted BIC, Scaling Correction Factor for MLR, and Entropy) for the Science Test (Paper-Based and Electronic Versions)

Test	No. of Classes	No. of Parameters	AIC	BIC	Adjusted BIC	Scaling Correction Factor (MLR)	Entropy
Science (Paper)	3	122	71526.3	72366.5	71916.0	1.04	0.80
	4	163	71568.2	72434.3	71979.5	1.08	0.75
	5	204	71718.0	72609.9	71961.9	1.14	0.72
Science (Electronic)	3	122	71344.36	71917.57	71605.0	1.07	0.83
	4	163	70996.00	71992.05	71399.2	1.09	0.80
	5	204	71050.32	72079.21	71431.2	1.08	0.81

From Table 7, it is observed that for the paper-based science test, the three-class latent model is the most appropriate according to the information criteria (BIC, AIC, Scaling Correction Factor for MLR, and Entropy). For the electronic science test, two models are acceptable:

- The three-class model, supported by BIC, Scaling Correction Factor for MLR, and Entropy
- The four-class model, supported by AIC and Adjusted BIC

Therefore, it was necessary to rely on classification probabilities, as shown in Table 8.

Table 8: Classification Probabilities for Most Likely Latent Class Membership by Class

Test	No. of Classes	No. of Parameters	Class 1	Class 2	Class 3	Class 4	Class 5
Science (Paper)	3	122	0.891	0.915	0.928		
	4	163	0.897	0.823	0.874	0.839	
	5	204	0.681	0.782	0.864	0.806	0.907
Science (Electronic)	3	122	0.929	0.920	0.890		
	4	163	0.873	0.899	0.868	0.863	
	5	204	0.999	0.906	0.870	0.897	0.858

The classification probabilities should be highest for each respective latent class. From Table 8, it is observed that in the paper-based science test, the probabilities are high across all three classes in the three-class model, while the probabilities are low across all four classes in the four-class model, and four probabilities are low in the five-class model. In the electronic science test, the probabilities are high in all three classes in the three-class model, while all four probabilities are low in the four-class model, and three probabilities are low in the five-class model.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Conclusion on the Number of Latent Performance Levels

Based on all results, the number of latent performance levels among eighth-grade students on the National Science Test is three latent classes, as follows:

Paper-Based Science Test:

- Lower Latent Class: Students who scored less than 14
  - Mean score: 11
- Middle Latent Class: Students who scored between 14 and 21 (cut score: 14)
  - Mean score: 17
- Upper Latent Class: Students who scored between 22 and 40 (cut score: 22)
  - Mean score: 27

Electronic Science Test:

- Lower Latent Class: Students who scored less than 15
  - Mean score: 12
- Middle Latent Class: Students who scored between 15 and 22 (cut score: 15)
  - Mean score: 18
- Upper Latent Class: Students who scored between 23 and 40 (cut score: 23)
  - Mean score: 27

To determine an individual's membership in a latent class in the model, the posterior conditional probability must be calculated. This probability must be at least 0.5, indicating the difference between conditional and unconditional membership. Therefore, the Posterior Conditional Probabilities were calculated to determine the classification probabilities, and accordingly, to define the membership of each student in the study sample within the respective latent class of the science test.

Table 9: Number and Percentage of Students and Membership Probability in Latent Classes for the Science Test

Test	Latent Class	Count	Percentage	Probability of Membership in Class 1	Class 2	Class 3
Science (Paper)	Class 1	492	32.8%	0.899	0.101	0.000
	Class 2	770	51.3%	0.070	0.907	0.022
	Class 3	238	15.9%	0.000	0.063	0.937
	Total	**1500**	**100%**	**32.8%**	**51.33%**	**15.87%**
Science (Electronic)	Class 1	816	54.4%	0.908	0.033	0.060
	Class 2	333	22.2%	0.080	0.920	0.000
	Class 3	351	23.4%	0.085	0.000	0.915
	Total	**1500**	**100%**	**54.4%**	**22.2%**	**23.4%**

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

From Table 9, we observe the posterior (conditional) membership probabilities for the paper-based science test:

- Class 3, consisting of 238 students (15.87% of the sample), ranks first in classification reliability based on its posterior conditional probability of 0.937.
- Class 2, with 770 students (51.33%), ranks second, with a posterior probability of 0.907.
- Class 1, with 492 students (32.8%), ranks third, with a posterior probability of 0.899.

In terms of unconditional membership (based on class size):

- Class 2 has the highest proportion of students (51.33%, 770 students),
- Followed by Class 1 (32.8%, 492 students),
- Then Class 3 (15.87%, 238 students).

For the electronic science test, the posterior conditional membership probabilities indicate:

- Class 2, with 333 students (22.2% of the sample), ranks first, with a posterior probability of 0.920.
- Class 3, with 351 students (23.4%), ranks second, with a probability of 0.915.
- Class 1, with 816 students (54.4%), ranks third, with a probability of 0.908.

Regarding unconditional membership based on class proportions:

- Class 1 ranks first, including 54.4% of the students (816 students),
- Followed by Class 3 at 23.4% (351 students),
- Then Class 2 at 22.2% (333 students).

Second: What are the probabilities of students answering each science test item correctly across the different latent performance levels?

Table 10 below shows the probabilities of students answering each science test item correctly across the different latent performance levels.

Table 10: Probabilities of Students Answering Each Science Test Item Correctly Across Latent Performance Levels

Item No.	Paper-Based Science Test	Electronic Science Test
	Lower Latent Class	Middle Latent Class
(1)	0.291	0.425
(2)	0.602	0.401
(3)	0.254	0.361
...	...	...
(40)	0.169	0.338

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Observations from Table 10:

Paper-Based Science Test:

- Students in the upper latent class had a probability of correctly answering 36 items above the 0.50 threshold, except for items 11, 12, 22, and 27, where their probabilities were below 0.50.

  This indicates that high-performing students belong to the upper latent class.

- Students in the lower latent class had a probability above 0.50 on only three items: items 2, 4, and 10.

- Students in the middle latent class had a probability above 0.50 on seven items: items 13, 15, 19, 20, 21, 23, and 30.

  This suggests that moderate-performing students belong to the middle latent class, and low-performing students belong to the lower latent class.

Electronic Science Test:

- Students in the upper latent class had a probability above 0.50 on 33 items, except for items 7, 11, 12, 22, 27, 34, and 35, where their probabilities were below 0.50. This again confirms that high-performing students are classified into the upper latent class.

- Students in the middle latent class had a probability above 0.50 on 13 items: items 4, 6, 9, 10, 13, 14, 15, 19, 20, 21, 23, 28, and 30.

- Students in the lower latent class had a probability above 0.50 on only two items: items 4 and 10.

This supports that moderate-performing students are in the middle latent class, and low-performing students belong to the lower latent class.

DISCUSSION OF RESULTS AND RECOMMENDATIONS

First: Discussion of Results Related to Research Question One

The results revealed the presence of three latent classes, based on the BLRT (Bootstrapped Likelihood Ratio Test), which is considered one of the best statistical indicators for identifying the appropriate number of latent classes.

Additional support came from the VLMR and LMR tests, as well as the information criteria: BIC, AIC, Adjusted BIC, Scaling Correction Factor for MLR, and Entropy. Based on these, individuals were classified into three classes: lower, middle, and upper. There were significant differences in responses between the lower and upper classes across most test items, as well as between the middle and upper classes, and between the middle and lower classes. To determine which items contributed to classifying students into the three latent classes, the Odds Ratios were calculated across the three classes for the science test.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

In the paper-based version, 27 items showed statistically significant differences in responses between the middle and lower latent classes. The remaining 13 items showed similar response patterns: items 11, 12, 16, 22, 24, 25, 27, 32, 33, 34, 37, 38, and 39.

Students in the middle and upper classes showed similar responses on items 10 and 22, and differed on the remaining 38 items. Meanwhile, students in the lower and upper classes differed on 39 items, with only item 35 showing similar responses. In the electronic version of the science test, 24 items showed significant differences in responses between the middle and lower classes, while 16 items had similar responses: items 1, 3, 5, 8, 11, 16, 22, 24, 25, 29, 31, 33, 34, 37, 38, and 39.

The middle and upper classes responded similarly on three items: 10, 27, and 35, and differed on 37 items. Meanwhile, students in the lower and upper classes differed on 35 items, and shared similar responses on five: 10, 12, 13, 22, and 27. These findings suggest that the key reason for differentiation among student abilities into latent classes lies in the prevalence of errors in inquiry skills, which were more common than errors in classification and tracking processes and life cycles.

This indicates a clear weakness in the foundational skill of knowledge acquisition. If a student lacks competence at this foundational level, their performance at subsequent levels (application and analysis) will also be poor.

This result—higher error rates in knowledge-level items compared to those measuring application and analysis—may be attributed to teachers relying heavily on teacher-centered, lecture-based strategies, and infrequently using modern, student-centered approaches that require active student engagement. As a result, students tend to lack motivation to memorize or understand information, leading to more errors at the knowledge level than at higher cognitive levels.

Additional factors may include the teacher's academic and professional preparation, the curriculum design in terms of the nature, difficulty, and structure of content, and limited school resources and infrastructure in Jordan—factors that significantly hinder the implementation of modern teaching practices.

The analysis ultimately yielded three latent classes, which does not align with the findings of Brown (2007), whose study of seventh- and eighth-grade students in mathematics revealed two latent classes. Nor does it align with Jarar and Bani Ata (2018), who found two latent classes for the algebra content domain, influenced by student gender, school location, and whether the teacher provided problem-solving explanations. Additionally, the results diverge from Sideridis et al. (2021), whose study aimed to identify student achievement levels based on demographic characteristics, parental background, and school behaviors such as absenteeism.

Their results revealed the presence of four latent classes, based on BIC, Bayes Information Criterion, and several other model fit indices proposed by Masyn (2013). In that study, parental education and student absenteeism significantly influenced latent class membership as positive and negative predictors of academic achievement.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

Second: Discussion of Results Related to Research Question Two

The results related to the paper-based science test indicated clear differences in the probability values of answering each item correctly.

Upon analyzing the data, it was found that all items effectively distinguished between the upper and middle classes, and that 97.5% of the items clearly distinguished between the upper and lower classes, with higher probabilities of correct responses in the upper class than in the other two classes.

Additionally, 80% of the items distinguished between the middle and lower classes, where the probability of correct responses was higher in the middle class than in the lower class. This indicates that the three latent classes are clearly distinguishable, with the lower class characterized by notable weakness in science, while the upper class demonstrates strong mastery of science content. Similarly, the results for the electronic science test revealed differences in the probability values across items.

Analysis showed that 95% of the items clearly distinguished between the upper and middle classes, and 97.5% between the upper and lower classes, with higher probabilities of correct answers in the upper class. Furthermore, 87.5% of the items distinguished between the middle and lower classes, with the middle class showing higher probabilities of correct responses. This also supports the conclusion that the three latent classes are clearly differentiated, with the lower class showing notable weaknesses in science, and the upper class demonstrating clear mastery of the subject.

These findings are consistent with the study by Cogo-Moreira et al. (2013), where the researchers applied Latent Class Analysis (LCA) to reading and writing subtest results and identified three latent classes that demonstrated strong discriminative power and explanatory capability.

They concluded that experimental methods such as LCA are effective in achieving accurate classification. The current study also aligns with the findings of Bani Ata (2022), who used LCA and found three latent classes, with a significant relationship between these classes and students' academic degree levels.

Recommendations:

1. There is a **clear weakness in student performance** on the science test, and the **reasons behind this weakness are not well understood**. Therefore, there is a need to investigate whether this underperformance reflects an actual deficiency in content mastery.
2. It is necessary to **reconsider the cut scores** currently adopted by the Ministry of Education and to apply a **recognized standard-setting method** instead of relying on **arbitrary approaches**.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

References

1) Bani Ata, Z. S. I. (2022). *Using latent class analysis to identify thinking styles of education faculty students based on Harrison and Bramson's theory*. The Educational Journal, 36(142), 97–125. Retrieved from http://search.mandumah.com/Record/1243911

2) Jarrar, N. A., & Bani Ata, Z. S. (2018). *Latent classes of eighth-grade students' performance on TIMSS mathematics tests in Jordan*. Educational Sciences, 26(3), 328–380. Retrieved from http://search.mandumah.com/Record/1087910

3) Allam, S. D. M. (2007). *Criterion-referenced diagnostic tests in educational, psychological, and training fields*. Cairo: Dar Al-Fikr Al-Arabi.

4) Ministry of Education. (2022). *Results of the national test for monitoring the quality of education for the academic year 2021–2022 for eighth-grade students*. Directorate of Examinations.

5) المراجع الأجنبية:

6) Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley and Sons, New York. 325–331. https://doi.org/10.1002/9780470114759.ch11

7) Angoff, W. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike, (Ed.), *Educational Measurement*. Washington, DC: American Council in Education.

8) Berk, R. (1986). A consumer's guide to setting performance standards on Criterion referenced tests. *Review of Educational Research*, 56(1), 137-172.

9) Binici, S. and Cuhadar, I. (2022). Validating Performance Standards via Latent Class Analysis. *Journal of Educational Measurement*, 59: 502- 516. https://doi.org/10.1111/jedm.12325

10) Brown, R. (2007). Using Latent Class Analysis to Set Academic Performance Standards. *Educational Assessment*, 12, 283–301. https://doi.org/10.1080/10627190701578321

11) Byrd, C. & Andrews, D. J. (2016). Variations in students' perceived reasons for, sources of, and forms of in-school discrimination: A latent class analysis. *Journal of School Psychology*, *57*, 1–14. https://doi.org/10.1016/j.jsp.2016.05.001

12) Cogo-Moreira, H., Carvalho, C., De Souza Batista Kida, A., De Ávila, C. R. B., Salum, G. A., Moriyama, T. S., Gadelha, A., Rohde, L. A., Moura, L. M., Jackowski, A. P., & De Jesus Mari, J. (2013). Latent class analysis of reading, decoding, and writing performance using the Academic Performance Test: concurrent and discriminating validity. *Neuropsychiatric Disease and Treatment*, 1175. https://doi.org/10.2147/ndt.s45785

13) Deanna, L. & Garbo, R. (2005). A Dynamic and Functional Approach as a Basis for Individual Educational Planning in Inclusive Contexts*. Erdelyi Pszichologiai Szemle, SI.* https://hdl.handle.net/10281/29735

14) Ebel, R. (1972). *Essentials of educational measurement*. Englewood Cliffs.

15) Hagenaars, J. & McCutcheon, A. (2002). *Applied latent class analysis* (pp. 89–106). *Cambridge, UK: Cambridge University Press*.

16) Halpin, G., Sigmon, G., & Halpin, G. (1983). Minimum competency standards set by three divergent groups of raters using three judgmental procedures: Implications for validity. *Educational and Psychological Measurement, 43*(1), 185–196. https://doi.org/10.1177/001316448304300126

17) Hambleton, R. (1978). Contributions to Criterion-Referenced Testing Technology: An Introduction. *Review of Educational Research*. 48, 223- 249.

18) Hambleton, R. (1982). Advances in criterion-referenced testing technology. In C.R. Reynolds &T. Gutkin (eds.), *Handbook of school psychology*. New York: Wiley, pp. 351-379.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

19) Jaeger, R. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis*, 4, 461-476.

20) Karakoyun, F., & Başaran, B. (2021). Identifying Turkish students' profiles of using information and communication technologies and its relationship with their academic achievement: A latent class analysis approach. *E-learning and Digital Media*, *19*(3), 295–319. https://doi.org/10.1177/20427530211060919

21) Killian, M., Cimino, A., Weller, B., & Hyun Seo, C. (2019). A systematic review of latent variable mixture modeling research in social work journals. *Journal of Evidence-Based Social Work*, 16(2), 192-210. https://doi.org/10.1080 /23761407.2019.1577783

22) Li, R., Zhou, W., & Wu, J. (2021). Identifying the subtypes of psychological profiles in senior undergraduate nursing students and its relationship with academic performance: A latent class analysis. *Journal of Professional Nursing*, *37*(4), 757–764. https://doi.org/10.1016/j.profnurs.2021.04.011

23) Linn, R., Koretz, D., & Baker, E. (1996). *Assessing the validity of the National Assessment of Educational Progress*: NAEP Technical Review Panel white paper (CSE Tech. Rep. No. 416). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

24) Lin, J. (2005). *The Bookmark Standard Setting Procedure: Strengths and Weaknesses*. The center for research in Applied Measurement and Evaluation. The university of Alberta.

25) Livingston, S. & Zieky, M. (1982). *Passing scores*: *A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.

26) Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: a comparison with K-means. *Canadian Journal of Marketing Research*, 20(1), 36–43.

27) McMullen, J., Lewis, R. W., & Bailey, D. (2020). Latent classes from complex assessments: What do they tell us? *Learning and Individual Differences*, 83–84, 101944. https://doi.org/10.1016/j.lindif.2020.101944

28) Muthén, L. & Muthén, B. (2017). *Mplus user's guide (8th ed.)*. Muthén & Muthén. Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.

29) Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, *78*(6), 1706–1722. https://doi.org/10.1111/j.1467-8624.2007.01097.x

30) Reckase, M. (2001). Innovative methods for helping standard-setting participants to perform their task: The role of feedback regarding consistency, accuracy, and impact. In: G. J. Cizek (Ed.), *Setting performance standards: Concepts methods and perspectives* (pp. 159-173). *Mahwah, NJ: Erlbaum.* https://doi.org/10.3403/00567985u

31) Rose, T., Lindsey, M., Xiao, Y., Finigan-Carr, N., & Joe, S. (2017). Mental health and educational experiences among Black Youth: A Latent class analysis. *Journal of Youth and Adolescence*, *46*(11), 2321–2340. https://doi.org/10.1007/s10964-017-0723-3

32) Sideridis, G., Tsaousis, I., & Al-Harbi, K. (2021). Identifying Student Subgroups as a Function of School Level Attributes: A Multilevel Latent Class Analysis. *Frontiers in Psychology*, 12, 231 – 250. https://doi.org/10.3389/fpsyg.2021.624221

33) Simpson, E. (1951). The interpretation of interactions in contingency tables. *Journal of the Royal Statistical Society*, B (13), 238–241. https://doi.org/10.1177/019263655103518201

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 58 Issue: 04:2025
DOI: 10.5281/zenodo.15281726

34) Sireci, S. & Biskin, B. (1992). Measurement practices in national licensing examination programs: A survey. *Clear Exam Review*, 3, 21-25.

35) Vaval, L., Bowers, A. & Rangel, V. (2019). Identifying a typology of high schools based on their orientation toward STEM: A latent class analysis of HSLS:09. *Science Education*, *103*(5), 1151–1175. https://doi.org/10.1002/sce.21534

36) Vermunt, J. K., & Magidson, J. (2000). *Latent GOLD's user's guide*. Boston: Statistical Innovations, Inc.

37) Vermunt, J. & Magidson, J. (2004). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20, 37–44.

38) Wyse, A. & Babcock, B. (2017). An Investigation of Undefined Cut Scores with the Hofstee Standard-Setting Method. *Educational Measurement: Issues and Practice*. 36. 10.1111/emip.12163.

39) Yang, X., Shaftel, J., Glasnapp, D., & Poggio, J. (2005). Qualitative or Quantitative Differences? *Journal of Special Education*, 38(4), 194–207. https://doi.org/10.1177/00224669050380040101