International Journal of Contemporary Educational Research (IJCER)

www.ijcer.net

Comparison of International TIMSS 2011 Proficiency Levels and Cut-off Scores Set by Using Cluster Analysis

Mahmut Sami Koyuncu[1], Ayşenur Erdemir[1]
[1]Gazi University

To cite this article:

Koyuncu, M.S. & Erdemir, A. (2019). Comparison of international TIMSS 2011 proficiency levels and cut-off scores set by using cluster analysis. *International Journal of Contemporary Educational Research, 6*(1), 155-168.
DOI: https://doi.org/10.33200/ijcer.544194

International Journal of Contemporary Educational Research
Volume 6, Number 1, June 2019, 155-168 ISSN: 2148-3868

Comparison of International TIMSS 2011 Proficiency Levels and Cut-off Scores Set by Using Cluster Analysis*

Mahmut Sami Koyuncu[1], Ayşenur Erdemir[1]**
[1]Gazi University

Abstract

The aim of the study is to compare TIMSS 2011 proficiency levels with the proficiency levels defined by the researchers using cluster analysis for Turkey, Korean, Norway, and Morocco in 4th and 8th grades in the fields of science and mathematics. Therefore, it is tried to be reached that these cut-off scores for each country can serve the evaluation of each country itself. For this research, the data gathered from related countries' students was taken from TIMSS 2011 database. Statistical analysis was performed with SPSS Version 21.0 statistic software package. The cut-off scores for these four countries selected in this study for each grade level and course type were defined using cluster analysis. Then, proficiency levels according to these cut-off scores were compared to the general TIMSS 2011 proficiency levels, and so the difference between these levels and percentage of agreement have been examined. According to the results, cut-off scores set by using cluster analysis for Korea were higher than TIMSS international benchmarks. Cut-off scores set for Morocco, Norway, and Turkey were lower than TIMSS international benchmarks. the percentage of agreement of the proficiency levels was found to be between 8.1% and 70.0%, and in general, it has been found that the percentage of agreement was low. Consequently, it is suggested that countries should make a standard-setting study for their own samples instead of using TIMSS international benchmarks for their own evaluations.
Keywords: Standard Setting, Cluster Analysis, Validity, TIMSS 2011

Introduction

Trends in International Mathematics and Science Study (TIMSS) is a survey study on the knowledge and skills of 4th and 8th-grade students in mathematics and science. The general aim of TIMSS is to measure student achievement in mathematics and science in the countries participating in the research, to determine how education and training take place in schools, the effectiveness and efficiency of the education system, and the differences between education systems of countries.

In order to assess student achievement in mathematics and science in TIMSS, in addition to the scores obtained by the students from the relevant test, the proficiencies related to these scores and the international benchmarks for these proficiencies were defined. In this way, countries can examine the proficiency level of their students and compare them with other countries. International benchmarks are defined by standard-setting study.

Standard-setting is essential for determining the differentiation in the success or performance levels of individuals. It is a standard setting study to determine whether the students are at a minimum level of proficiency or which point can be used as a cut-off point in order to be adequate in an area. (Berberoglu, 2009). Cizek (2001) describes the standard setting as the determination of performance levels for deciding or classifying individuals, while Crocker and Algina (1986) define it as specifying the cut-off point. The determined cut-off points are used to determine the qualifications of the students in specific areas, and they are also used to evaluate the performance of students in international exams such as TIMSS, PISA.

---

*A preliminary version of this paper was presented at the 26th International Conference on Educational Sciences that was held in Antalya, Turkey on April 20-23, 2017

** Corresponding Author: *Mahmut Sami Koyuncu*, *ms_koyuncu@hotmail.com*

In order to determine the proficiency levels in TIMSS scale anchoring method is used. Scale anchoring is a method that has a statistical component and uses item characteristics to discriminate points on the proficiency scale. Also, it has a consensus component, which means it involves educational experts. These experts and identified items are used to interpret which students at or close to the determined scale points. (Beaton and Allen, 1992). The procedures of this method applied for the first time in TIMSS 1999 were first included in the National Assessment of Educational Progress (NAEP) (Beaton and Allen, 1992; Gregory and Mullis,1999). Table 1 includes TIMSS international benchmarks used to explain the performance of the students in the test items and the level of proficiency corresponding to these criteria. Besides, the level of proficiency of students who are below 400 points not mentioned in the table is expressed as the below-low international benchmark (Mullis, Martin, Foy & Arora, 2012).

Table 1. TIMSS 2015 International Benchmark

Scale Score	International Benchmark
625	Advanced International Benchmark
550	High International Benchmark
475	Intermediate International Benchmark
400	Low International Benchmark

TIMSS international proficiency levels are used to compare countries with each other. However, there are no criteria for evaluating countries within themselves. For this reason, it is thought to be necessary to set the new cut-off points that will allow the countries to be evaluated within themselves. In this study, k-means method was used for clustering analysis methods to determine the cut-off points that will be used in the evaluation of the countries. Cluster analysis that can be used as an alternative standard setting method is the process of separating individuals or items into groups, called clusters, by using information from a set of data according to specific proximity criteria. In the clustering process, the similarity of the elements in the cluster should be high, whereas the similarity between the clusters should be low. In cluster analysis, the k-means algorithm is commonly used and is the best-known algorithm. The k-means algorithm is used to separate the given objects into k-classes according to their attributes or characteristics. The algorithm is called k-means because a fixed number of clusters is needed before the algorithm runs. The k represents the number of sets and also the number of groups to occur. Accordingly, k is a constant positive integer which is known in advance and does not change its value until the clustering process is finished. The classification in the k-means is carried out by placing the data around the cluster centers (centroid) to which they are the closest or similar (Dinçer, 2006).

Standard setting studies with cluster analysis are not common. Sireci, Robin, and Patelis (1999) compared the standards obtained by cluster analysis with the standards determined by the more traditional methods which are borderline surveys and contrasting groups. Violato, Marini, and Lee (2003) examined the validity of the expert judgment by comparing the minimum performance levels determined by the Nedelsky and Ebel methods based on the expert judgment of the certification examinations with the cut-off scores determined by the cluster analysis. Hess, Subhiyah, and Giordano (2007) investigated the effectiveness of cluster analysis to verify the minimum pass scores determined by the Angoff method.

A review of the relevant literature reveals that there is no study in which the TIMSS international benchmarks are compared with those of defined by different standard setting methods. Studies on TIMSS international benchmarks were generally based on the determination of international criteria for student achievement. (Gregory,1999; Gregory & Mullis, 1999; Kelly, Mullis & Martin, 2000; Kelly, 2002; Mullis, Erberber & Preuschoff, 2007). Ker (2013) compared the TIMSS 2011 mathematics achievement of the Chinese Taipei, Singapore, and the USA countries at International Benchmark levels. Olsen and Nilsen (2017) examined the standard setting methods used in TIMMS and PISA and discussed how these procedures could be used locally in tests and evaluations in the Norwegian context.

In addition, the cut-off scores for TIMSS international proficiency levels allow countries to compare each other while there are no cut-off points for the countries to be evaluated within themselves. For this reason, it is evident that there is a lack of focus of the validity of the standard-setting procedures and the cut-off points for the evaluation of each country independently and there is a need to work on this issue.

The aim of the study is to compare TIMSS 2011 proficiency levels with the proficiency levels defined by the researchers using cluster analysis for Turkey, Korean, Norway, and Morocco in 4^{th} and 8^{th} grades in the fields of science and mathematics. Therefore, it is tried to be reached that these cut-off scores for each country can serve the evaluation of each country independently. In this context, the research questions are as follows:

1. Do TIMSS international benchmark and cut-off point determined by cluster analysis differ according to grade level and course type?
2. Do TIMSS international benchmarks and cut-off points obtained from clustering analysis vary according to the percentage of agreement between proficiency levels according to the grade level and type of the course?
3. Do the percentage of the students in the proficiency levels obtained by the cluster analysis and the percentage of the students in the proficiency levels determined according to the TIMSS international proficiency levels differ according to the grade level and type of course?

Method

Research design

Since the aim of the study is to compare TIMSS 2011 international benchmarks with the proficiency levels defined by the researchers using cluster analysis for Turkey, Korean, Norway and Morocco in 4th and 8th grades in the fields of science and mathematics, type of the study can be defined as descriptive research.

Study Group

In TIMSS 2011, 42 countries at the 8th grade and 50 countries at the 4th grade participated. The study group was determined by taking into account the success ranking of the countries and the participation in the TIMSS 2011 at both grades. In this context, Korea having high achievement for both mathematics and science in each grade level, Norway having moderate achievement, Morocco having low achievement and Turkey just for the comparison have been selected for the study group. Table 2 shows the numbers and rates of the 8th grade and 4th-grade students of the four countries included in the research.

Table 2. TIMSS 2011 Descriptive statistics for 4th and 8th-grade students

Country	8th grade		4th grade	
	f	%	f	%
Morocco	8986	36.0	7841	34.4
Turkey	6928	27.8	7479	32.8
Korea	5166	20.7	4334	19.0
Norway	3862	15.5	3121	13.7
Total	24942	100.0	22775	100.0

As indicated in Table 2, the highest number of students in 4th-grade and 8th-grade is in Morocco with low performance, and the least participation is in Norway with moderate performance.

The mean scores of science and mathematics at the 8th-grade level of the four countries constituting the study group and the rankings of them among all countries are given in Table 3.

Table 3. TIMSS 2011 The rankings and mean scores for 8th grade

Country	8th-grade mathematics		8th-grade science	
	Ranking	Mean score	Ranking	Mean score
Korea	1	613	3	560
Norway	20	475	19	494
Turkey	24	452	21	483
Morocco	40	371	41	376

Korea's 8th grade mathematics achievement test mean score is 613 and its ranking is 1; Norway's mean score is 475, and its ranking is 20; Turkey's mean score is 452, and its ranking is 24, and the mean score of Morocco is 371 and its ranking is 40, which can be found in Table 3. Korea's 8th-grade science achievement test mean score is 560 and its ranking is 3; Norway's mean score is 494 and its ranking is 19; Turkey's mean score is 483 and its ranking is 21, and the mean score of Morocco is 376 and its ranking is 41.

The mean scores of science and mathematics at the 4th-grade level of the four countries of the study group and the achievement rankings among all countries are given in Table 4.

Table 4. TIMSS 2011 The rankings and mean scores for 4th grade

Country	4th-grade mathematics		4th-grade science	
	Ranking	Mean score	Ranking	Mean score
Korea	2	605	1	587
Norway	29	495	33	494
Turkey	35	469	36	463
Morocco	49	335	49	264

Korea's 4th grade mathematics achievement test mean score is 605 and its achievement ranking is 2; Norway's mean score is 495, and its ranking is 29; Turkey's mean score is 469, and its ranking is 35, and the mean score of Morocco is 335, and its ranking is 49, which can be found in Table 4. Korea's 4th-grade mathematics achievement test mean score is 587 and its achievement ranking is 1; Norway's mean score is 494, and its ranking is 33; Turkey's mean score is 463, and its ranking is 36, and the mean score of Morocco is 264, and its ranking is 49.

Data

TIMSS offers an international comparison of student achievements and provides countries with information on curricula. TIMSS conducted every four years by the International Association for the Evaluation of Educational Achievement (IEA) also constitute an international database to specify students' trends in mathematics and science achievement.

For this research, the data gathered from the selected countries' (Morocco, Korea, Norway, and Turkey) students were taken from TIMSS 2011 database. The data are available from the TIMSS 2011 international database (http://timss.bc.edu/timss2011/international-database.html).

Data Analysis

Before the primary data analysis is run, data screening was done. As a result of data screening, it was determined that there was no missing data in TIMSS datasets and there were no extreme values that would affect the analysis of the results. Statistical analyses were performed with SPSS software package. Cut-off scores for these four countries selected in this study for each grade level and course type were defined using k-means clustering. The reason for using the k-means method is that the number of groups, k = 5, is known in advance. According to TIMSS international benchmarks, there are four achievement benchmarks which are low, intermediate, high, and advanced.

In the cluster analysis, the average of the plausible values in TIMSS for each grade and course type (8th-grade mathematics BSMMAT01-05, science BSSSCI01-05; 4th-grade mathematics ASMMAT01-05; science ASSSCI01-05) was used as the achievement score of mathematics and science. Principle components analysis was performed to obtain a single variable from 5 plausible values in TIMSS. As a result of this analysis, factor scores of single factor structure were obtained by the regression method. Factor scores were found to be the same as z scores obtained from the mean of 5 plausible value. Therefore, this study was performed on the average of 5 plausible values. Also, the analysis was carried out on a single variable as it was aimed to determine the cut-off point as well as the cluster of the students. Since a single variable was used, no conversion was made for the clustering analysis.

The average of five plausible values is defined as student achievement score, and the students are divided into five groups by cluster analysis. Then, the minimum and maximum test scores within each cluster were computed. The average of the maximum score of the low group and the minimum score of the high group was used as the cut score. In this study, according to each grade level and course type, the cut-off points were determined by cluster analysis. The proficiency levels defined by the researchers were compared with TIMSS international benchmarks in terms of the percentages of students meeting the specified benchmarks.

Findings and Interpretation

Do TIMSS international benchmark and cut-off points determined by cluster analysis vary according to grade level and course type?

Table 5 shows the minimum and maximum scores of the proficiency levels determined by the cluster analysis for the 8th-grade mathematics achievement.

Table 5. TIMSS 2011 8th-grade mathematics achievement proficiency levels

8th-grade mathematics achievement proficiency levels										
COUNTRY	Below low (below 400)		Low (400 - 475)		Intermediate (475 - 550)		High (550 - 625)		Advanced (at or above 625)	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
KOREA	281.29	489.03	489.19	570.62	570.88	638.69	638.74	703.89	704.10	850.59
NORWAY	203.19	360.78	361.47	414.65	414.70	468.07	468.18	526.07	526.17	647.05
TURKEY	142.85	334.65	334.77	421.54	421.60	506.97	507.09	600.73	600.74	844.40
MOROCCO	127.99	307.39	307.43	373.42	373.46	438.86	438.94	511.19	511.40	666.53

The average of the minimum and maximum values in Table 5 is determined as the cut-off point. Figure 1 shows the cut-off points for the 8th-grade mathematics achievement.

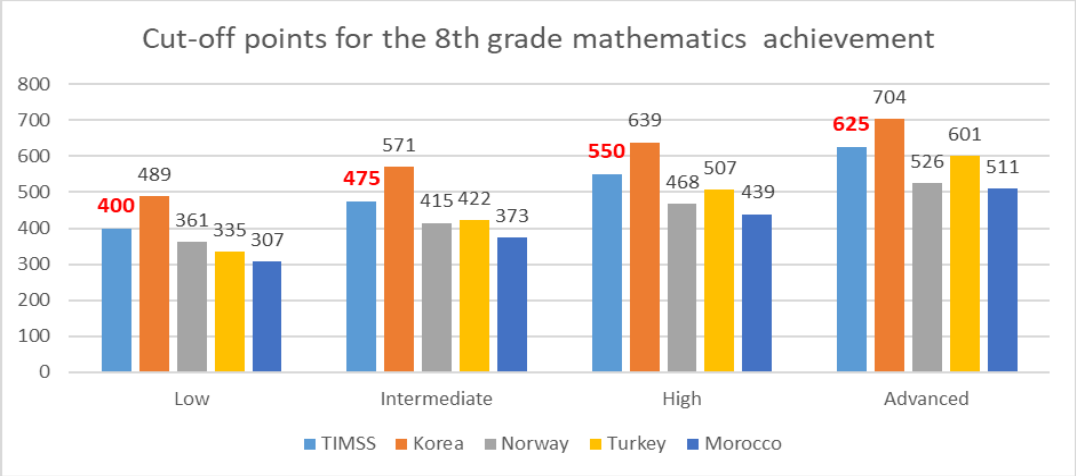


Figure 1. TIMSS 2011 cut-off points for the 8th-grade mathematics achievement

Only the cut-off scores determined by the cluster analysis for Korea are higher than the international benchmarks in TIMSS 2011 while those for Norway, Turkey, and Morocco are lower than the international benchmarks, which can be found in Figure 1.

The country with the highest deviation from the TIMSS international benchmarks is Morocco whose ranking is lower than the others. The cut-off scores calculated for Turkey with cluster analysis for 8th-grade mathematics achievement are the closest to the TIMSS international benchmarks except for low benchmark. The nearest cut-off point to the low benchmark belongs to Norway.

Table 6 shows the minimum and maximum scores of the proficiency levels determined by the cluster analysis for the 8th-grade science achievement.

Table 6. TIMSS 2011 8th-grade science achievement proficiency levels

8th-grade science achievement proficiency levels										
COUNTRY	Below low (below 400)		Low (400 - 475)		Intermediate (475 - 550)		High (550 - 625)		Advanced (at or above 625)	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
KOREA	294.77	456.93	457.05	525.68	525.74	583.71	583.76	642.26	642.43	777.65
NORWAY	211.67	381.42	381.77	445.95	445.98	503.43	503.51	563.36	563.56	686.81
TURKEY	132.94	363.47	363.54	449.47	449.51	527.29	527.34	607.74	607.83	821.97
MOROCCO	107.56	307.26	307.40	375.73	375.77	439.66	439.71	505.94	506.15	681.27

Figure 6 shows the cut-off points determined by cluster analysis using minimum and maximum values for the 8th-grade science achievement in Table 6.

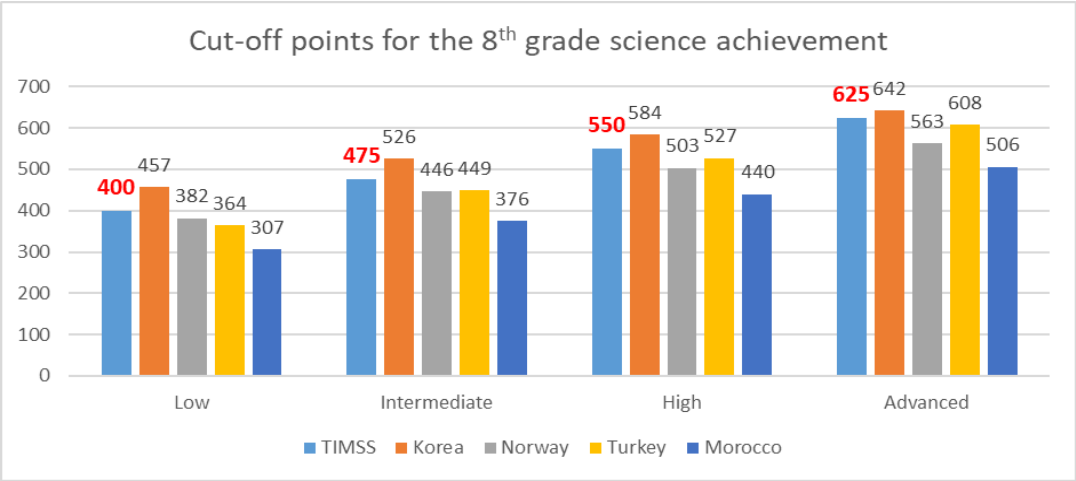


Figure 2. TIMSS 2011 cut-off points for the 8th-grade science achievement

Only the cut-off scores determined by the cluster analysis for Korea are higher than the international benchmarks in TIMSS 2011 while those for Norway, Turkey, and Morocco are lower than the international benchmarks, which can be found in Figure 2.

The country with the highest deviation from the TIMSS international benchmarks is Morocco whose ranking is lower than the others. In general, cut-off scores calculated for Turkey with cluster analysis for 8th-grade science achievement are the closest to the TIMSS international benchmarks except for low benchmark. The nearest cut-off point to the low benchmark belongs to Norway.

Table 7 shows the minimum and maximum scores of the proficiency levels determined by the cluster analysis for the 4th-grade mathematics achievement.

Table 7. TIMSS 2011 4th-grade mathematics achievement proficiency levels

4th-grade mathematics achievement proficiency levels										
COUNTRY	Below low (below 400)		Low (400 - 475)		Intermediate (475 - 550)		High (550 - 625)		Advanced (at or above 625)	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
KOREA	308.51	481.78	482.70	546.37	546.60	601.68	601.74	658.56	658.66	784.09
NORWAY	259.62	399.24	399.27	455.43	455.47	507.81	508.13	565.06	565.17	698.82
TURKEY	119.79	342.46	342.57	431.47	431.50	503.74	503.82	574.97	575.04	756.64
MOROCCO	113.79	266.32	266.39	340.38	340.40	415.94	416.16	496.99	497.14	680.85

Figure 3 shows the cut-off points determined by cluster analysis using minimum and maximum values for the 4th-grade math achievement in Table 7.

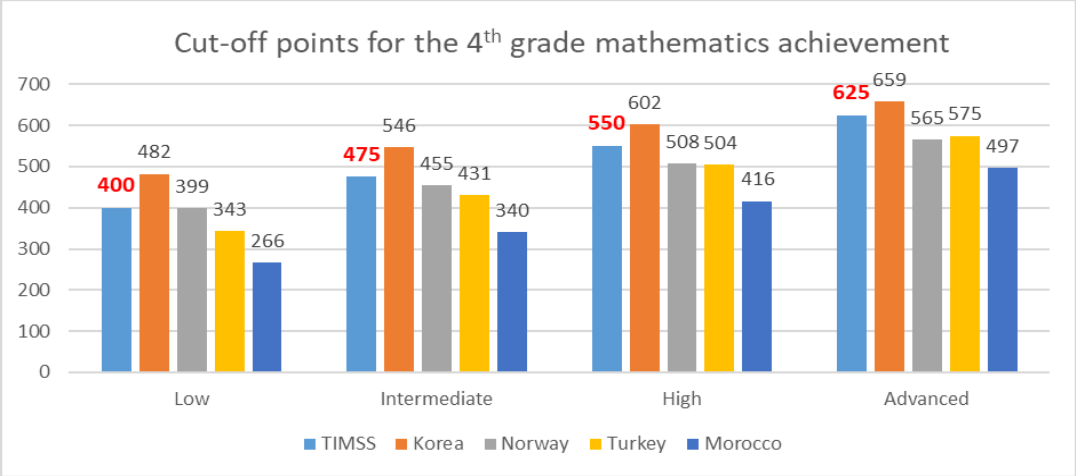


Figure 3. TIMSS 2011 cut-off points for the 4th-grade mathematics achievement

Only the cut-off scores determined by the cluster analysis for Korea are higher than the international benchmarks in TIMSS 2011 while those for Norway, Turkey, and Morocco are lower than the international benchmarks, which can be found in Figure 3.

The country with the highest deviation from the TIMSS international benchmarks is Morocco whose ranking is lower than the others. In general, cut-off scores calculated for Norway with cluster analysis for 4th-grade mathematics achievement are the closest to the TIMSS international benchmarks except for high benchmark. The nearest cut-off point to the low benchmark belongs to Turkey.

Table 8 shows the minimum and maximum scores of the proficiency levels determined by the cluster analysis for the 4th-grade science achievement.

Table 8. TIMSS 2011 4th-grade science achievement proficiency levels

4th-grade science achievement proficiency levels										
COUNTRY	Below low (below 400)		Low (400 - 475)		Intermediate (475 - 550)		High (550 - 625)		Advanced (at or above 625)	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
KOREA	318.15	484.25	484.98	547.72	547.77	598.80	598.90	650.03	650.09	761.34
NORWAY	264.20	422.71	422.76	478.94	479.07	525.19	525.33	569.67	570.15	691.42
TURKEY	116.10	333.75	334.41	422.57	422.63	494.18	494.19	565.56	565.70	756.02
MOROCCO	24.87	164.30	164.40	246.62	246.71	333.29	333.45	433.31	433.59	677.01

Figure 4 shows the cut-off points determined by cluster analysis using minimum and maximum values for the 4th-grade science achievement in Table 8.

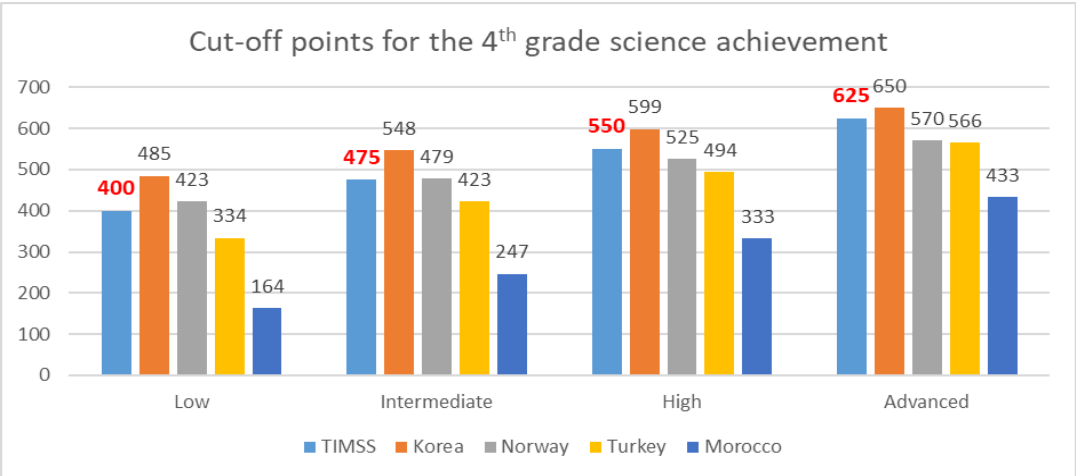


Figure 4. TIMSS 2011 cut-off points for the 4th-grade science achievement

Only the cut-off scores determined by the cluster analysis for Korea are higher than the international benchmarks in TIMSS 2011 while those for Norway, Turkey, and Morocco are lower than the international benchmarks, which can be found in Figure 4.

The country with the highest deviation from the TIMSS international benchmarks is Morocco whose ranking is lower than the others. In general, cut-off scores calculated for Norway with cluster analysis for 4th-grade science achievement are the closest to the TIMSS international benchmarks.

Do TIMSS international benchmarks and cut-off points obtained from cluster analysis differ according to the percentage of agreement between proficiency levels according to the grade level and type of the course?

Figure 5 shows the percentage of agreement for students having the same level of proficiency determined by cluster analysis with four international proficiency levels of TIMSS 2011. For example, according to the 4th-grade mathematics achievement 37.1% of students in Korea are in the same proficiency level in both proficiency levels determined by clustering analysis and TIMSS international benchmarks.

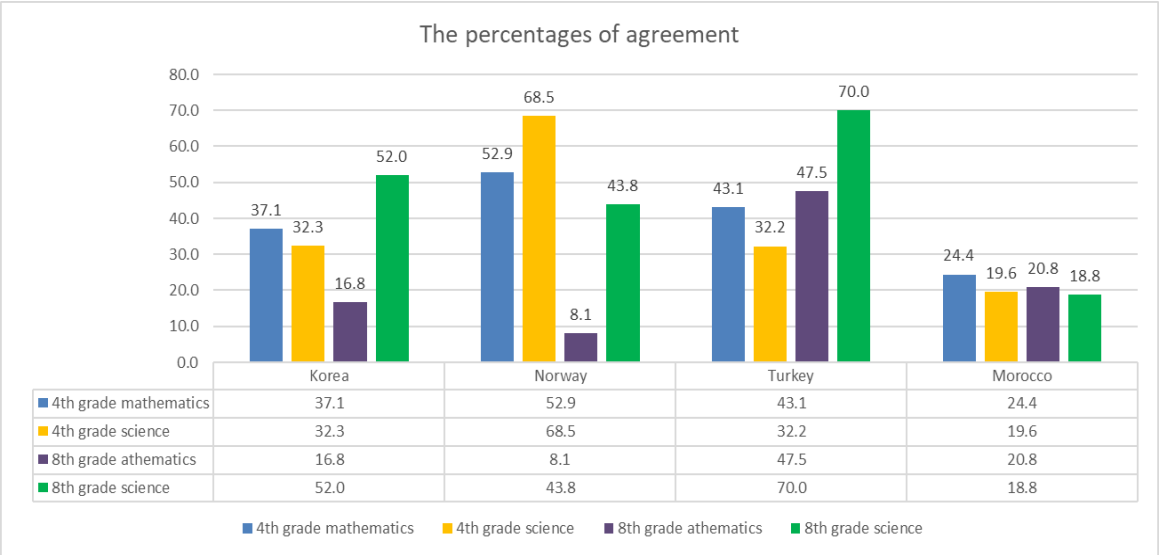


Figure 5. TIMSS 2011 the percentages of agreement

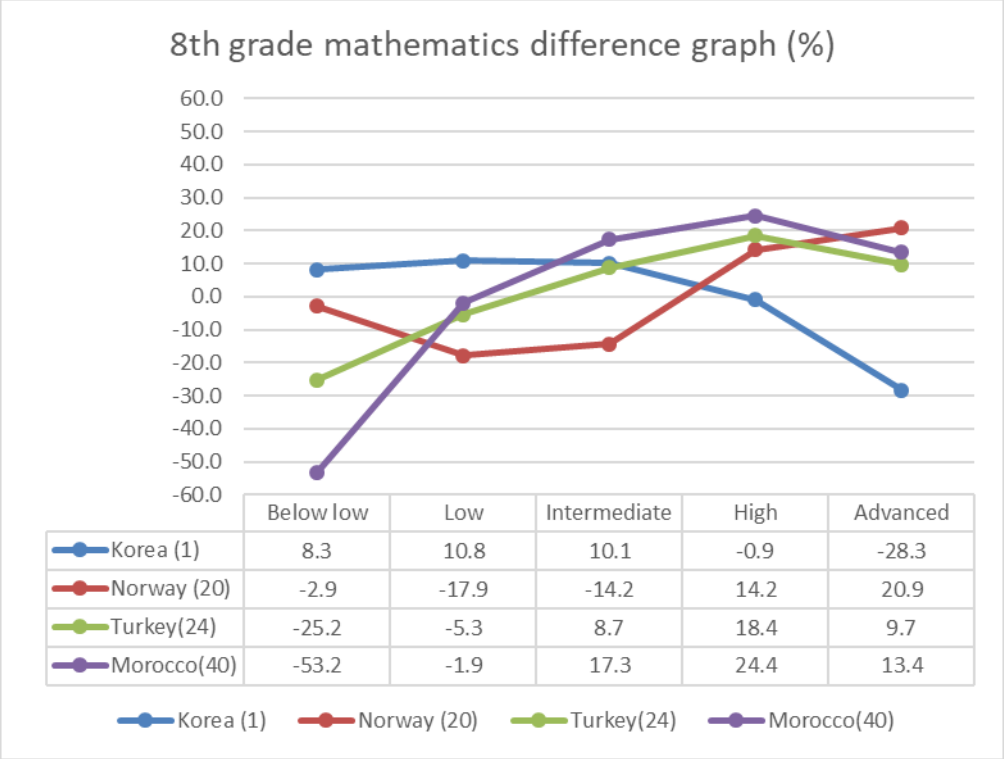
The percentage of agreement varies between 16.8 - 52.0% for Korea; 8.1 - 68.5 for Norway; 32.2 - 70.0% for Turkey, and 18.8 - 24.4% for Morocco, which can be found in Figure 5.

When the percentages of agreement of the proficiency levels are examined by grade and type of course, the highest percentage of agreement for TIMSS 2011 4th grade mathematics achievement is between TIMSS international benchmarks and proficiency levels determined by cluster analysis for Norway (52.9%) and the lowest is between TIMSS benchmarks and proficiency levels determined by cluster analysis for Morocco (24.4%). The highest percentage of agreement for TIMSS 2011 4th grade science achievement is between TIMSS international benchmarks and proficiency levels determined by cluster analysis for Norway (68.5%) and the lowest is between TIMSS benchmarks and proficiency levels determined by cluster analysis for Morocco (19.6%). On the other hand, for8th-grade mathematics achievement the highest percentage of agreement is between TIMSS international benchmarks and proficiency levels determined by cluster analysis for Turkey (47.5%) and the lowest is between TIMSS benchmarks and proficiency levels determined by cluster analysis for Norway (8.1%). For 8th grade science achievement, the percentage of agreement between TIMSS international benchmarks and cluster analysis is the highest for Turkey (70.0%) and the lowest for Norway (8.1%). The percentage of agreement is generally low.

Do the percentage of the students in the proficiency levels obtained by the cluster analysis and the percentage of the students in the proficiency levels determined according to the TIMSS international proficiency levels differ according to the grade level and type of course?

For the 8th grade mathematics achievement, the difference between the percentages of the students in the proficiency levels determined by the cluster analysis and the percentages of the students in the international

proficiency levels of TIMSS 2011 is shown in Figure 6. Increases and decreases in the graphs are made concerning TIMSS 2011 international proficiency levels.



	Below low	Low	Intermediate	High	Advanced
Korea (1)	8.3	10.8	10.1	-0.9	-28.3
Norway (20)	-2.9	-17.9	-14.2	14.2	20.9
Turkey(24)	-25.2	-5.3	8.7	18.4	9.7
Morocco(40)	-53.2	-1.9	17.3	24.4	13.4

Figure 6. TIMSS 2011 8th-grade mathematics difference graph

For Morocco, the percentage of students that are at below-low level determined by cluster analysis for TIMSS 2011 8th-grade math achievement is 53.2% less than the percentage of students at the same level determined by TIMSS international benchmarks. It is 25.2% less for Turkey and 2.9% less for Norway. On the other hand, the percentage of students that are at below-low level determined by cluster analysis is 8,3% more than the percentage of students at the same level determined by TIMSS international benchmarks for Korea.

The percentage of students at low level decreases by 1.9% for Morocco; 17.9% for Norway and 5.3% for Turkey while it shows a 10.8% increase for Korea. The percentage of students at intermediate level shows an increase of 17.3% for Morocco, 10.1% for Korea, and 8.7% for Turkey. It decreases by 14.2 for Norway. The maximum increase in the percentage of students at high level is for Morocco with a percentage of 24.4%, which is followed by Turkey (18.4%) and Norway (14.2%). It decreases just for Korea with a percentage of 0.9%. Likewise, the percentage of students at advanced level shows an increase of 20.9% for Norway, 13.4% for Morocco, and 9.7% for Korea while it shows a decrease of 28.3% for Korea.

In general, the percentage of below-low-level and low-level students determined by TIMSS international benchmarks decreased in all three countries except Korea, while the percentage of high and advanced-level students increased in all countries except Korea. While there is a decrease in the percentage of intermediate-level students only in Norway, there is an increase in other countries.

For the 8th-grade science achievement, the difference between the percentages of the students in the proficiency levels determined by the cluster analysis and the percentages of the students in the international proficiency levels of TIMSS 2011 is shown in Figure 7.

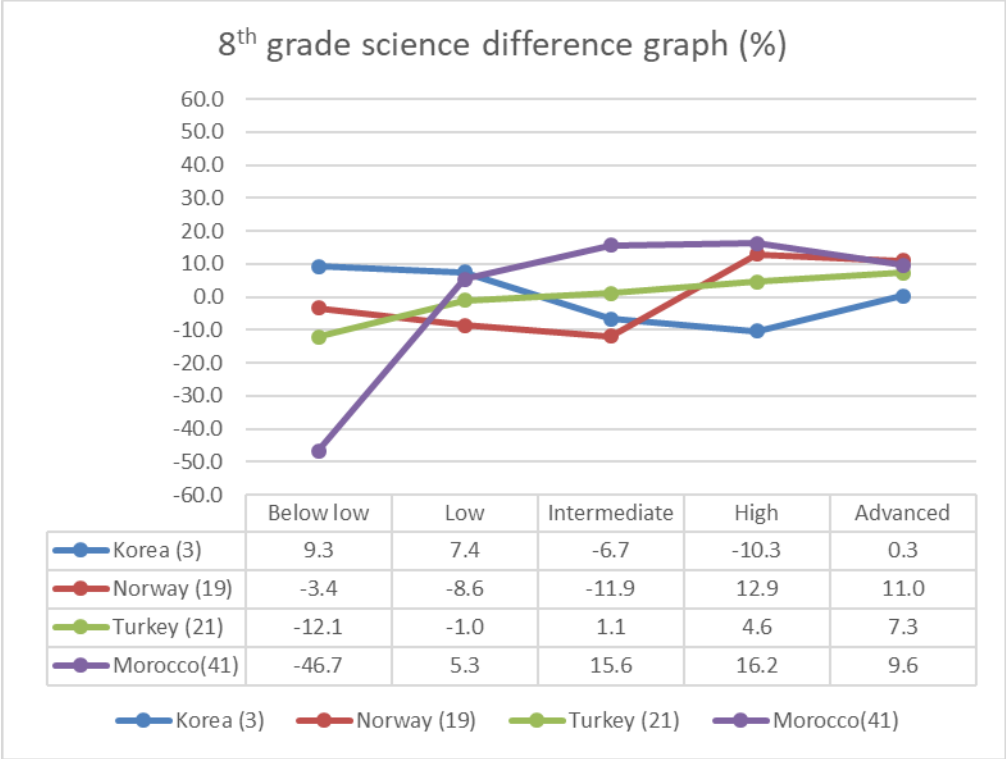


Figure 7. TIMSS 2011 8th-grade science difference graph

For Morocco, the percentage of students that are at below-low level determined by cluster analysis for TIMSS 2011 8th-grade science achievement is 46.7% less than the percentage of students at the same level determined by TIMSS international benchmarks. It is 12.1% less for Turkey and 3.4% less for Norway. On the other hand, the percentage of students that are at below-low level determined by cluster analysis is 9,3% more than the percentage of students at the same level determined by TIMSS international benchmarks for Korea.

The percentage of students at low level decreases by 8.6% for Norway and 1.0% for Turkey while it shows a 7.4% increase for Korea and 5.3% for Morocco. The percentage of students at intermediate level shows an increase of 15.6% for Morocco and 1.1% Turkey. However, it shows a decrease of 11.9% for Norway and 6.7% for Korea. The maximum increase in the percentage of students at high level is for Morocco with a percentage of 16.2%, which is followed by Norway (12.9%) and Turkey (4.6%). It decreases just for Korea with a percentage of 10.3%. Likewise, the percentage of students at advanced level shows an increase of 11.0% for Norway, 9.6% for Morocco, 7.3% for Turkey, and 0.3% for Korea.

In general, the percentage of below-low-level students determined by TIMSS international benchmarks decreased in all three countries except Korea while the percentage of high-level students increased in all countries except Korea. While there is not much difference in the percentage of advanced-level students in Norway, there is an increase in other countries.

For the 4th-grade mathematics achievement, the difference between the percentages of the students in the proficiency levels determined by the cluster analysis and the percentages of the students in the international proficiency levels of TIMSS 2011 is shown in Figure 8.

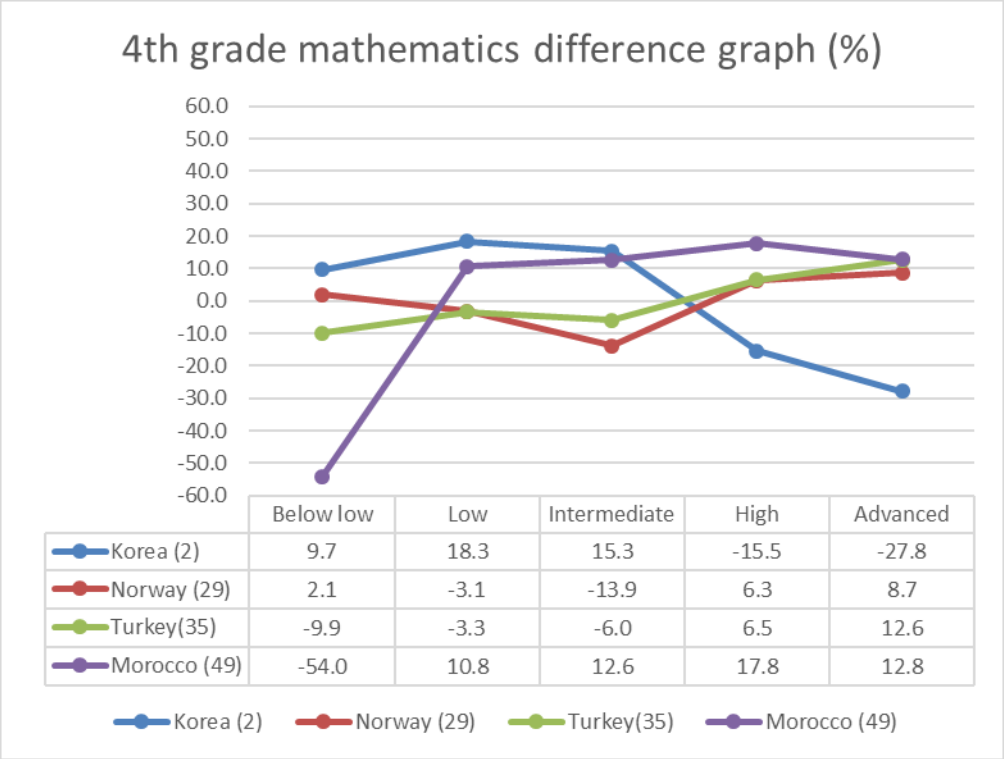


Figure 8. TIMSS 2011 4th-grade mathematics difference graph

For Morocco, the percentage of students that are at below-low level determined by cluster analysis for TIMSS 2011 4th-grade mathematics achievement is 54.0% less than the percentage of students at the same level determined by TIMSS international benchmarks. It is 9.4% less for Turkey. On the other hand, the percentage of students that are at below-low level determined by cluster analysis is 9,7% more than the percentage of students at the same level determined by TIMSS international benchmarks for Korea and 2.1% for Turkey.

The percentage of students at low level decreases by 3.3% for Turkey and 3.1% for Norway while it shows an 18.3% increase for Korea and 10.8% for Morocco. The percentage of students at intermediate level shows an increase of 15.3% for Korea and 16.2% for Morocco. However, it shows a decrease of 13.9% for Norway and 6.0% for Turkey. The maximum increase in the percentage of students at high level is for Morocco with a percentage of 17.8%, which is followed by Turkey (6.5%) and Norway (6.3%). It decreases just for Korea with a percentage of 15.5%. Likewise, the percentage of students at advanced level shows an increase of 12.8% for Morocco, 12.6% for Turkey, and 8.7% for Norway while it decreases by 27.8% for Korea.

In general, the percentage of below-low-level students determined by TIMSS international benchmarks decreased in Turkey and Morocco, while the percentage of high-level and advanced-level students increased in all countries except Korea.
For the 4th-grade science achievement, the difference between the percentages of the students in the proficiency levels determined by the cluster analysis and the percentages of the students in the international proficiency levels of TIMSS 2011 is shown in Figure 9.

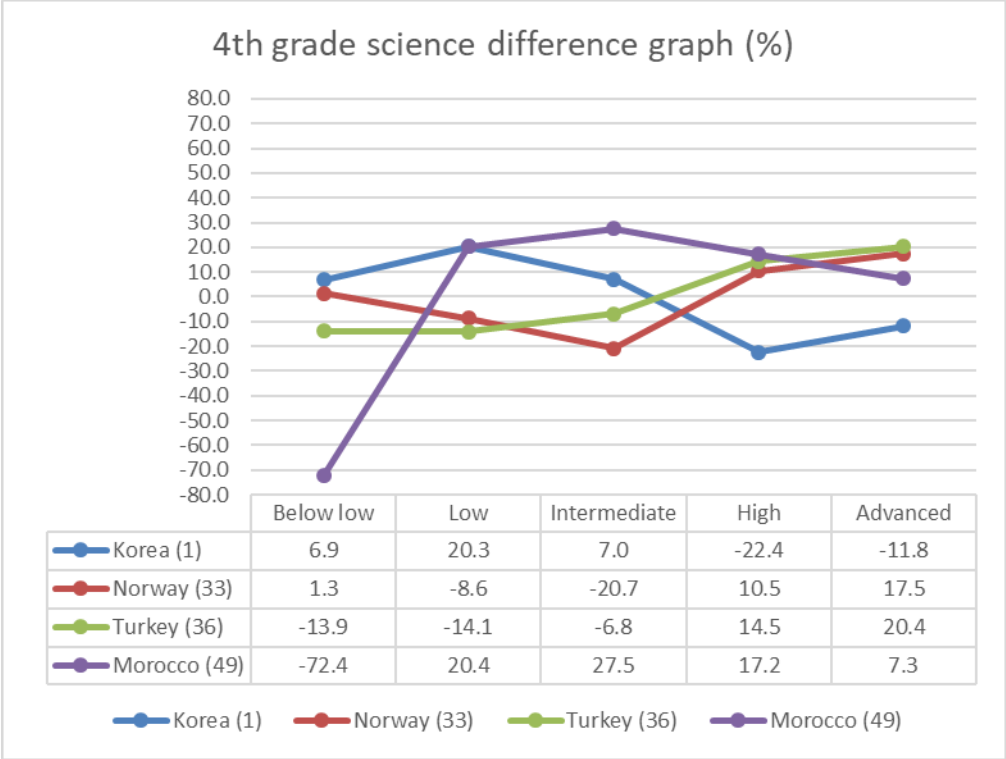


Figure 9. TIMSS 2011 4th-grade science difference graph

For Morocco and Turkey, the percentage of students that are at below-low level determined by cluster analysis for TIMSS 2011 4th-grade science achievement is less than the percentage of students at the same level determined by TIMSS international benchmarks, 72.4% and 13.9% respectively. On the other hand, the percentage of students that are at below-low level determined by cluster analysis is 6.9% more than the percentage of students at the same level determined by TIMSS international benchmarks for Korea and 1.3% for Norway.

The percentage of students at low-level decreases by 14.1% for Turkey and 8.6% for Norway while it shows an increase of 20.3% for Korea and 20.4% for Morocco. The percentage of students at intermediate-level shows an increase of 27.5% for Morocco and 7.0% for Korea. However, it shows a decrease of 20.7% for Norway and 6.8% for Turkey. The maximum increase in the percentage of students at high-level is for Morocco with a percentage of 17.2%, which is followed by Turkey (14.5%) and Norway (10.5%). It decreases just for Korea with a percentage of 22.4%. Likewise, the percentage of students at advanced level shows an increase of 20.4% for Turkey, 17.5% for Norway, and 7.3% for Morocco while it decreases by 11.8% for Korea.

In general, the percentage of below-low-level students determined by TIMSS international benchmarks decreased in Turkey and Morocco while the percentage of high-level and advanced-level students increased in all countries except Korea.

Conclusion

The aim of the study is to compare TIMSS 2011 proficiency levels with the proficiency levels defined by the researchers using cluster analysis for Turkey, Korea, Norway and Morocco in 4th and 8th grades in the fields of science and mathematics. Therefore, it is tried to be achieved that these cut-off scores for each country can serve the evaluation of each country itself. According to the results, cut-off scores set using cluster analysis for Korea were higher than TIMSS international cut-off scores in both grade levels and both of the fields, mathematics and science. Cut-off scores set for Morocco, Norway, and Turkey were lower than TIMSS international cut-off scores.

Morocco, which has a low success rate among all the countries, has the most significant difference in cut-off scores between its own cut-off scores set using cluster analysis and the international ones. Nonetheless, the lowest differences between cut-off scores are in Turkey for 8th grades and Norway for 4th grades. When the

agreement between international proficiency levels and the ones defined using cluster analysis was examined, the highest percentage of agreement was obtained for Turkey and Norway, and the lowest percentage of agreement was obtained for Korea and Morocco. It has been found that the percentage of agreement was low overall.

For TIMSS 2011 4th and 8th grade mathematics and science achievement, the percentage of below-low-level students determined by TIMSS international benchmarks decreases in Turkey, Norway, and Morocco while it increases in Korea. The situation is the opposite for high and advanced level determined by clustering analysis. The percentage of high and advanced-level students determined by TIMSS international benchmarks increases in Turkey, Norway, and Morocco, while it decreases in Korea. So, depending on their proficiency level determined by cluster analysis, the number of students at below-low-level in Morocco, Norway, and Turkey decreases while it increases in Korea; The number of students at high and advanced-level in Morocco, Norway and Turkey increases while it decreases in Korea. As a result, depending on their proficiency levels determined by cluster analysis, Morocco, Norway and Turkey's successes are higher than the ones defined by the TIMSS international benchmarks; The success of Korea determined by cluster analysis is lower than TIMSS proficiency levels.

In general, when TIMSS international proficiency levels are compared with the proficiency levels obtained by cluster analysis, it can be said that TIMSS international benchmarks provide an advantage to successful countries (Korea) while they have a reverse situation for the other countries having low success. This is an expected situation because the TIMSS determines the level of proficiencies of all countries and compares the rankings of countries.

Recommendations

The cut-off points obtained by the cluster analysis allow the countries to evaluate themselves in the local sense. Consequently, it is suggested that countries should make a standard-setting study for their own samples instead of using TIMSS international benchmarks for their own evaluations. In studies using the TIMSS plausible value, the cut-off scores obtained in this study can be used to evaluate students' achievements. Similar studies can be carried out using other standard setting methods besides cluster analysis. In this study, four countries were selected for the analysis. The cut-off scores for the data of other countries can be set in the future.

References

Beaton, A. E., & Allen, N. L. (1992). Chapter 6: Interpreting Scales Through Scale Anchoring. *Journal of Educational Statistics*, 17(2), 191-204.

Berberoğlu, G. (2009). Madde Haritalama Yöntemi ve Cito Türkiye Öğrenci İzleme Sistemi (ÖİS) Uygulamalarında Yeterlik Düzeylerinin Belirlenmesi [Item Mapping Method and the Determination of Proficiency Level in the application of Turkey Cito Student Monitoring System]. *Cito Eğitim: Kuram ve Uygulama Dergisi*, Mayıs-Haziran 2009, 13-24.

Cizek, G. (2001). Conjectures on the rise and call of standard-setting: An introduction to context and practice. In G. J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 3–17). Mahwah, NJ: Erlbaum.

Crocker, L. ve Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: CBS.

Dinçer, Ş.E. (2006). *The K-Means Algorithm in Data Mining and an Application in Medicine.* Master's Thesis, Kocaeli University Insitute of Natural and Applied Sciences, Kocaeli.

Gregory, K. D. (1999). Describing TIMSS 1999 International Benchmarks of Student Achievement.

Gregory, K. D., & Mullis, I. V. S. (1999). Describing international benchmarks of student achievement. *TIMSS*, 267-278.

Gündeğer, C., Doğan, N. (June, 2014). *A Statistical Approach to Setting Standard: Cluster Analysis.* Paper presented at 4th International Conference on Measurement and Evaluation in Education and Psychology, Ankara, Turkey.

Hess, B., Subhiyah, R. G., & Giordano, C. (2007). "Convergence Between Cluster Analysis and the Angoff Method for Setting Minimum Passing Scores on Credentialing Examinations." Evaluation and the Health Professions, 30 (4), 362-375.

Kelly, D. L., Mullis, I. V., & Martin, M. O. (2000). *Profiles of student achievement in mathematics at the TIMSS international benchmarks: US performance and standards in an international context.* TIMSS International Study Center, Lynch School of Education, Boston College.

Kelly, D. L. (2002). The TIMSS 1995 international benchmarks of mathematics and science achievement: Profiles of world class performance at fourth and eighth grades. *Educational Research and Evaluation*, 8(1), 41-54.

Ker, H. W. (2013). Trend Analysis on Mathematics Achievements: A Comparative Study Using TIMSS Data. *Universal Journal of Educational Research*, 1(3), 200-203.

Mullis, I. V., Erberber, E., & Preuschoff, C. (2007). The TIMSS 2007 international benchmarks of student achievement in mathematics and science. *TIMSS*, 339-347.

Mullis, I. V., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 international results in mathematics*. International Association for the Evaluation of Educational Achievement. Herengracht 487, Amsterdam, 1017 BT, The Netherlands.

Olsen, R. V., & Nilsen, T. (2017). Standard Setting in PISA and TIMSS and How These Procedures Can Be Used Nationally. *In Standard Setting in Education* (pp. 69-84). Springer, Cham.

Sireci, S. G., Robin, F., & Patelis, T. (1999). Using cluster analysis to facilitate standard setting. *Applied Measurement in Education*, *12*(3), 301-325.

Violato, C., Marini, A., & Lee, C. (2003). A validity study of expert judgment procedures for setting cutoff scores on high-stakes credentialing examinations using cluster analysis. *Evaluation & the health professions*, *26*(1), 59-72.