

Examination of Student Growth Using Gain Score and Categorical Growth Models

Hatice Cigdem Yavuz ^{1,*}, Ömer Kutlu ²

¹Cukurova University, Faculty of Education, 01330, Sarıçam/Adana, Turkey

²Ankara University, Faculty of Educational Sciences, 06590 Çankaya/Ankara, Turkey

ARTICLE HISTORY

Received: 30 May 2019

Revised: 08 August 2019

Accepted: 31 August 2019

KEYWORDS

Gain score model,
Categorical growth model,
Student growth,
Multilevel modeling

Abstract: In this study, gain score, and categorical growth models were used to examine the role of student (gender and socioeconomic level) and school characteristics (school size and school resources) in the student growth on comprehension skills in language. The participants of this study were 2,416 sixth-grade students in 2011 who became seventh-grade students in 2012. The data was collected through two achievement tests, student and school questionnaires. Two achievement tests were calibrated using the Rasch Model and were scaled using the concurrent estimation method. Moreover, the cut-off scores of these tests were determined by using the bookmark method. Students' growth was modelled with the gain score and categorical growth models. All data was analyzed using multilevel models. Results showed that some students did not achieve sufficient gains to advance to higher performance levels. Although some schools' average gains were higher, their performance was still not significant enough in terms of tests' standards. Moreover, the analyses demonstrated that the student gain scores and growth categories varied significantly among the schools. In addition, the study was able to determine student and school characteristics that have an impact on the students' gain scores and categorical growth. Given the different aspects gained about students' performance with these models, it is recommended to utilize different growth models in schools.

1. INTRODUCTION

The widespread use of assessments in education which focus on students' performances determined from a single time point is a point of contention in the field of assessment studies (Betebenner & Linn, 2009). The reason for this is because the information obtained from such assessments is limited. Within the scope of this limited information in question, multiple questions arise concerning the validity of classification of students, determining of their performance levels and considering students with learning difficulties, as well as inference from teachers and schools (Laird, 2008). Furthermore, together with these assessments, education shareholders are able to see the growth of students, and they may be able to discern whether the growth in question is in accordance with the standards as well (Yen, 2007; cited in Betebenner & Linn, 2009). In this sense, assessments that measure development can provide more clear

CONTACT: Hatice Cigdem YAVUZ ✉ hcyavuz@cu.edu.tr 📧 Cukurova University, Faculty of Education, 01330, Sarıçam/Adana, Turkey

ISSN-e: 2148-7456 /© IJATE 2019

information regarding school effectiveness and student achievement (Heck, 2006). Indeed, applications that measure growth have been effectively used in many educational systems (e.g. Assessment Agency, 2008; NCLB [No Child Left Behind], 2002; U.S. Department of Education, 2010) for many years and are ever increasing in their importance (Briggs & Betebenner, 2009).

Studies focusing on student growth in the literature predominately concern themselves with comprehension skills in language and reading comprehension skills (Herbers, Cutuli, Supkoff, Heistad, Chan, Hinz, & Masten, 2012; Hughes, Luo, Kwo, & Loyd, 2008; McCoach, O'Connell, Reis, & Levitt, 2006; Skibbe, Connor, Morrison, & Jewkes, 2010). The reason for this is that reading comprehension and language skills are one of the most important fields in terms of educational accountability (Shin, Davison, Long, Chan, & Heistad, 2013). Moreover, these skills can play a significant role in student academic achievement in other subjects (Arnold & Doctoroff, 2003; Crawford, Tindal, & Steiber, 2001). In this sense, academic achievement at higher grade levels of students who have difficulty in reading comprehension are also affected negatively by such deficiencies in upcoming grades (Crawford et al., 2001; Herbers et al., 2012). In this context, it can be stated that the linguistic skills are of great importance for students' academic lives.

As in academic achievement, gender differences are presented in the academic growth of students as well. In the related literature, it is indicated by many studies that the growth of female students in language and reading comprehension skills generally surpasses that of male students (Denton & West, 2002; Husain & Millimet, 2009; Kurdek & Sinclair, 2001). In particular, studies pointed out that the students' initial level of reading comprehension is different according to gender (Morgan, Farkas, & Wu, 2011). Another crucial characteristic determining student success is the fact that the socioeconomic level of the student also plays an important role in the academic growth (McCoach et al., 2006; Nese, Biancarosa, Anderson, Lai, Alonzo, & Tindal, 2012). According to Shin et al. (2013), students coming from lower socioeconomic backgrounds have shown lower academic achievement in the fields of comprehension skills in language and reading than other students. Similarly, it has been seen that the growth of the students coming from lower socioeconomic backgrounds is slower than the other students (Palardy, 2008).

Another area of interest is the role of school size and school resources which are among the school characteristics examined in student growth varies according to school types in different countries and different studies (Hanushek, 2006; Stevensen, 2006). According to studies in the literature, the effect of school size on student achievement may vary in degree or even its direction; however, it's also possible that such an effect may not be observed (Stevenson, 2006). The same situation is also observed on student growth (Heck, 2006; Palardy, 2008). In the literature, the effect of school resources on the student's academic achievement has been studied effectively for many years (Hanushek, 2006). Studies in the literature reveal different results about the effects of school resources on student achievement (Krueger & Lindahl, 2001). In the literature, some studies reveal the positive effect of school resources on student growth (Cheti & Birgitta, 2012; Palardy, 2008), whereas others have not found a significant effect (Glewwe, Hanushek, Humpage, & Ravina, 2011).

Only a limited number of studies that pursue data on the academic growth of students have looked (Ergin-Aydemir & Sünbül, 2016; Bursal, 2013; Yapar, 2014) at Turkey. These studies did not focus on students' linguistic skills. Of these studies, only Yapar (2014) conducted a study on student growth in English reading skills. Although student growth is not monitored in reading comprehension skills in Turkey, there are studies that examine students' status in this field according to student and school characteristics (Erman-Aslanoğlu and Kutlu, 2015; Kutlu, Yıldırım, Bilican, & Kumandaş, 2011; Güzle-Kayır & Erdoğan, 2015; Özer-Özkan & Doğan,

2013). The results of these studies show that student and school characteristics effect students' reading comprehension. Large-scale assessments conducted in Turkey (e.g.; ABIDE [Monitoring and Evaluation of Academic Skills], PISA [The Programme for International Student Assessment] and PIRLS [Progress in International Reading Literacy Study]) show that Turkish students do not excel in reading comprehension (Organization for Economic Cooperation and Development [OECD], 2014a, 2016; Mullis, Martin, Gonzalez, & Kennedy, 2001). The findings of these studies seem to agree that student performance in reading comprehension might vary based on the student and school characteristics. However, the role of school and student characteristics in Turkish students' growth cannot be determined because there are no studies regarding monitoring students in Turkey. Nevertheless, outside of Turkey many studies in the literature have monitored the growth of comprehension skills in language and reading comprehension skills of students in the context of different student and school characteristics (McCoach et al., 2006; Palardy, 2008; Skibbe et al., 2010; Shin et al., 2013).

Since growth can be measured according to gains and/or norms, it should be asked whether growth is intended to be measured based on performance standards or groups (Gong, 2004). The change in student performance in the gain score model can be seen with a calculation made by subtracting the score obtained in the previous years from the score obtained in the relevant year (Welch, Dunbar, & Rickels, 2016). In the categorical growth model, student growth is converted into the performance levels corresponding to the student's scores and inferences are made based on these performance levels. In the context of this study, the modeling of growth in comprehension skills in language over a year has been modeled according to the gain scores and the performance levels. By including student and school characteristics into these two different student growth models within the framework of educational accountability, this study aims to determine the effects of these characteristics on student growth.

In Turkey, there is a lack of sufficient relevant data on students' academic growth and the growth of different student groups. Thus, how the role of students' performance levels changed in one year, and the role of school and student characteristics in this change is not known. Assessments measuring students' performance, which are used instead of student growth models, do not enable the Turkish education system to grow. School shareholders are excluded from the educational accountability system since there is no data/information source in the Turkish education system for comparison based on certain standards of accountability (Nayır, 2013). In this sense, some researchers have concluded the practices related to accountability in the Turkish education system are insufficient (Türkoğlu & Aypay, 2015) and these practices are not informative for the shareholders.

In this study, gain score, and categorical growth models were used to examine the role of student (gender and socioeconomic level) and school characteristics (school size and school resources) in the student growth on comprehension skills in the language. In other words, this study aims to determine students' gain scores and at the same time, to monitor growth according to performance levels. In this context, answers to the following research questions are sought in the study: (i) What are the frequencies of schools in the gain scores and in growth categories?; (ii) What are the effects of the student gender and socioeconomic status as well as school characteristics such as school size and school resources on student gain scores?; (iii) What are the effects of the student gender and socioeconomic level and school characteristics such as school size and school resources on student growth categories?

2. METHOD

2.1. Study Group

The participants of this study are composed of 52 schools and 2146 students (52.21% female) which participated in the "Learning Level Research (LLR)" LLR-1 and LLR-3 (Ayrıl, Özdemir, & Sadıç, 2011) that Altındağ Guidance and Research Center carried over the 2011-

2012 and the 2012-2013 academic years during the spring term. The students included in the study were in the sixth grade in the 2011-2012 academic year in Altındağ district of Ankara, and the same students who continued as seventh graders in the 2012-2013 academic year.

2.2. Data Collection Tools

2.2.1. The achievement tests

Two achievement tests were applied in LLR 1 and LLR 3 projects. The test used in LLR 1 was applied as 18 multiple choice items with four choices in the 2010-2011 academic year while the other one in LLR 3 were applied as 54 multiple choice items with four choices in the 2011-2012 academic year. Together with two field experts, a total of five experts were employed in the development of the test which was applied in LLR 1. Together with four field experts, a total of 10 experts were employed in the development of the test, which was implemented in the LLR 3. The items, which were designed to measure comprehension skills in language, were placed in the achievement tests. In the development of the test, PISA, PIRLS studies and primary school curriculums were taken into consideration in determining the skills to be dealt with in the tests (Ayril et al., 2011). Before creating the final forms of the tests, a pilot study was conducted, and the items were revised based on item statistics analysis (for more information; see Ayril et al., 2011).

In this study, two new sub-tests were created by selecting items from the tests in the LLR 1 and LLR 3. The item selection of both tests was conducted based on the Rasch Model of the Item Response Theory (IRT) after checking all necessary assumptions. While selecting the specific items, attention was paid to the contents of the items, the item and item-fit statistics. In this context, 13 items were selected each from LLR 1 and LLR 3 while three of which were pseudo-common items.

For the newly created tests, the names NLLR1 and NLLR3 were used, respectively. After finalizing item selection and preparing new test forms, the performances of the students who participated in the LLR 1 and LLR 3 were re-estimated according to the answers they gave to the relevant items in NLLR1 and NLLR3. In this sense, new scores of the students were estimated within the scope of newly created test forms.

The KR-20 reliability coefficient was found to be .62 for both tests. The fact that there are a limited number of items in the tests may prevent the reliability to be higher. As shown in Figure 1, the information appears to be much greater in the theta range between -1.5 to $+1.5$ for the NLLR 1 and NLLR 3.

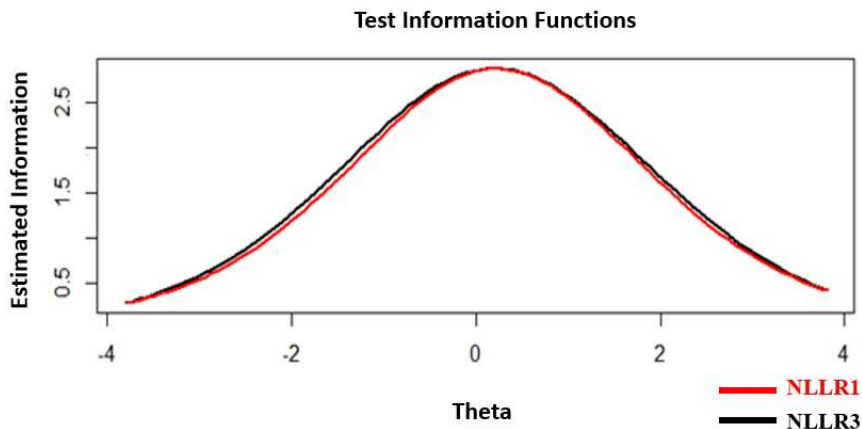


Figure 1. Test information functions for the NLLR 1 and NLLR 3

2.2.2. Student and School Questionnaires

In addition to the achievement tests, students were given questionnaires in LLR 1 and LLR 3. This study gathered information regarding the gender and socioeconomic level in the given

questionnaire. For the socioeconomic level of the students, an index variable was created by using the education level of the parents of the students, the number of books at home and the per capita income in the family. Moreover, the size of the school and the school resources (physical resources) were taken into consideration as school characteristics. In the data set, the school size variable was accepted as the total number of students. Another index variable was created for school resources in which the number of classrooms, the number of laboratories, music rooms, painting rooms, and the number of gyms were identified. The index calculation is explained in detail at the upcoming section.

2.3. Procedure and Data Analysis

Before conducting the data analysis, the missing values in the data set were checked. Firstly, data belonging to students who did not have any data at both of the two measurement points were excluded from the data set. This data set was used in statistical procedures during the preparation of the data for analysis. In the analyses conducted in order to find answers to the research questions, the data set excluding students without questionnaire data was used. In addition, since the parameter values could not be produced without bias in groups with clusters consisting of fewer than 12 units (Browne & Draper, 2006), students in schools with fewer than 12 students were excluded from the data set as well. Within this context, analyses were carried out on a total of 2004 students for the research questions. With the final data set, the data imputation was not done due to full information maximum likelihood (FIML) estimation used for the analyses.

The following steps were followed in preparing the data set for analysis:

1. Item calibration: The items in the achievement tests were calibrated according to the Rasch model for this study after checking all necessary assumptions (unidimensionality, local independence and item fit). Item difficulties ranged from $-.03$ to 1.53 for NLLR 1, and $.00$ to 1.53 for NLLR 3, with the average difficulty being $.00$ for both tests, which means that the items were of moderate difficulty overall.
2. Selection of the items: First of all, three pseudo-common items were chosen from both tests. The selection of pseudo-common items is based on Luppescu (2005)'s selection criteria regarding items to be used in virtual equating. In this sense, cognitive levels, subject areas and difficulty levels of the items were taken into consideration in the selection of pseudo-common items. After the selection of pseudo-common items with priority, the items were selected in a way that the number of items in the subject fields of the new tests to be created would be equal, and have similar reliability coefficients and average difficulties. According to this, initially, five items with low item quality were removed from LLR 1. Thus, with the 10 left and three pseudo-common items, NLLR 1 was created. After that, a total of 10 items were selected from the LLR 3 that were in accordance with the items in the NLLR 1. Thus, NLLR 3 test was created. For NLLR 1 and 3, item calibration based on the Rasch model was done again.
3. Scaling of tests: The tests were scaled so that the scores obtained from the tests used in this study would be interpreted correctly. In this study, the scaling of the tests was conducted by concurrent estimation, which is one of the IRT methods. In the concurrent estimation, all parameters can be on the same scale since all parameters of the items in both tests are estimated concurrently in a single run (Hanson & Beguin, 2002). For this reason, these methods do not require any conversion between forms (Gonzalez & Wiberg, 2017).
4. Converting scores from the ability parameters into test scores: As two performance levels were determined for the tests created in this study, the obtained ability parameters of the students were converted into scores in the range of 0-200.
5. Standard setting for the tests: In this study, the cut-off scores for NLLR 1 and 3 were determined, as tests with determined standards should be used in order to employ the categorical

growth models. The bookmark method, which is among test-centered bookmark methods (Lewis, Mitzel, & Green, 1996), was used for this research.

According to the bookmark method, two performance levels namely “basic level” and “proficient level” were determined for NLLR 1 and 3. In this study, six female Turkish-language teachers working in a primary school located in the Mamak district of Ankara served as panelists during the standard-setting process. The average year of seniority of the panelists was 8.83, but it ranged between 4 and 18 years.

In the study, before the standard setting of the tests, panelists underwent training related to the standard setting, the bookmark method and the tasks expected from. Considering the characteristics of the tests used in the study and the information that should be given for the training of the panelists, an agenda was formed for the standard setting panel. The panel was conducted in parallel with this agenda which was determined for the standard setting. At the end of three rounds with the panelists regarding the setting of the standards, the panelists determined the cut-off scores for NLLR 1 as 101.23 and 98.78 for NLLR 3.

6. Calculation of gain scores: After the tests were scaled, the achievement scores of the students in the year 2011 were subtracted from the achievement scores of 2012.

7. Determining the growth categories: Categorical growth models are defined as student growth which is converted into categorical performance levels corresponding to the student's scores, and making inferences over performance levels. In this context, according to the determined cut-off scores of NLLR 1 and 3, students' test scores were converted into the performance levels in both tests. Afterwards, the change students displayed from NLLR 1 to NLLR 3 was categorized. For this purpose, the students who pass towards a higher performance level were coded as 3, students who remained at a high-performance level were coded as 2, students remained at low-performance level are coded as 1 and students whose performance level downgraded to a basic level were coded as 0. In this sense, it should be noted that this categorization was made based on dummy coding, not on an ordinal categorization.

8. Creation of indexes of socioeconomic level and school resources: In the creation of the index variable, the index formula (OECD, 2014b, p. 352) was employed. In the calculation of the values in the formula, principal component analysis was used for the school resources variables; mixed principal component analysis was used for the socioeconomic level variables as those mentioned variables were composed of continuous and categorical variables together. Mixed principal component analysis was conducted in the R program with the package named as “PCAmixdata” (Chavent, Kuentz, Labenne, Liquet, & Saracco, 2014).

Descriptive statistics were used in the first research question of the study and multilevel models were used separately for the second and third research questions. In the study, the first level was taken as the student level and the second level was taken as the school level. There are two reasons for employing these models in the study. The first one is that the data set used in the study is nested, and these models can analyze more than one level in these structures more reliably (Raudenbush & Bryk, 2002). The other reason is to take into account that other analyses, apart from multilevel models require the assumption that the observations are independent. However, the students are not randomly placed in schools and the multilevel models are able to eliminate said problem (Osborne, 2000).

In the second research question, multilevel models (one-way ANOVA, first-level random intercept and regression model in which means are outcomes) were used. Mplus 8 program was used in the analysis of these models. The Mplus 8 program is advantageous and strong in terms of parameter prediction (Muthén & Muthén, 1998; 2017). In the third research question, the hierarchical generalized linear model (HGLM) was used. The HGLMs are different from multilevel models because the dependent variable does not meet the normality assumption and they are more compatible in terms of distributions (Raudenbush, Bryk, Cheong, & Congdon,

2011). The stated analyses were carried out with HLM 7.03 program (Raudenbush et al., 2011). The Pratt index was calculated in order to determine the importance of predictor variables in predicting the dependent variable which is taken with multilevel models. Pratt index is used to calculate the relative importance of predictor variables (Liu, Zumbo, & Wu, 2014).

3. FINDINGS

3.1. Findings of the models using gain score model

Descriptive statistics related to the values calculated according to the gain score model of schools are given in Appendix 1 and the results based on statistics are given in Figure 2.

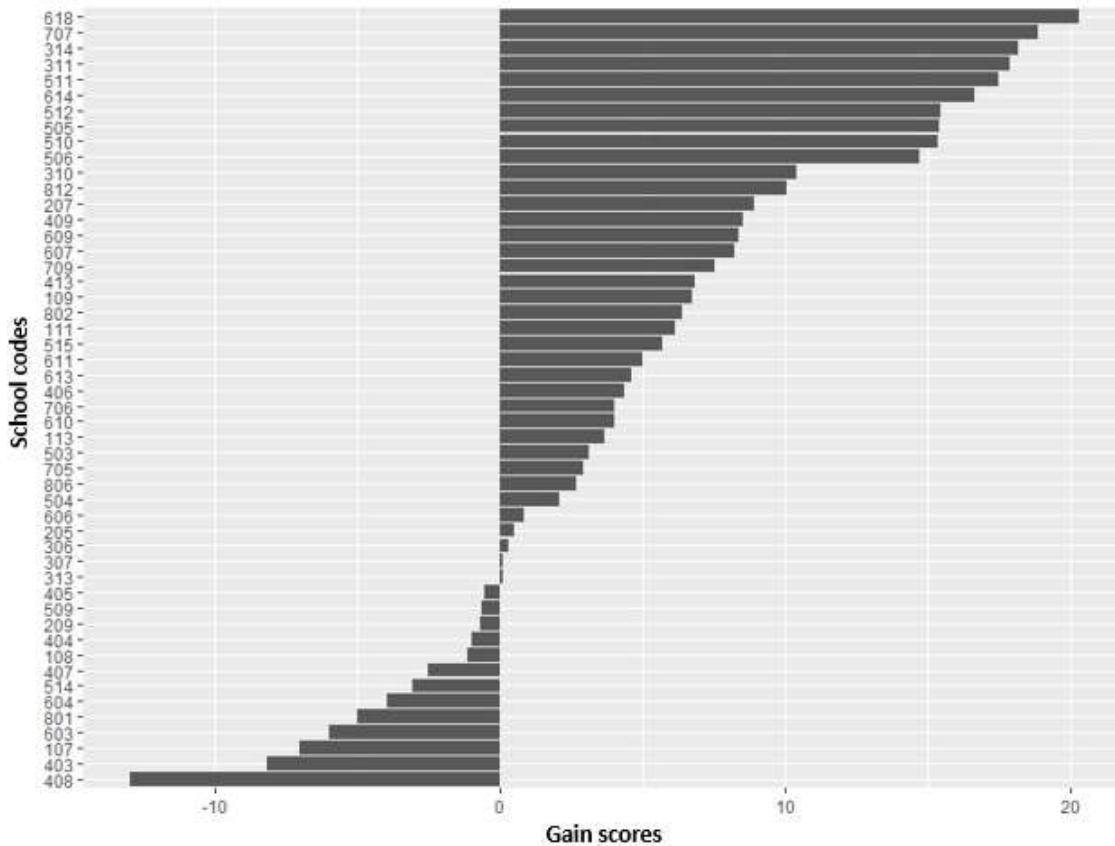


Figure 2. Frequency distribution of the average gain scores of schools

According to Figure 2, in terms of the gain model of the students who participated in the study, the school with the code of 618 has the highest average score, while the 408 coded school has the lowest average score. In addition, according to Appendix 1, the most homogenous school is the school with the code of 107, while the most heterogeneous school is the school with the code 707 in terms of gain scores.

3.2. Findings related to differences between schools in terms of gain scores

The results regarding the models established to examine the differences between schools in terms of the gain scores are given in Table 1.

Table 1. Results of random effects one-way ANOVA model

Variance components	Variance	df	Variance/df	p
Student level	1247.87	45.17	27.62	0.00
School level [u_{0j}]	30.99	11.39	2.72	0.01
Variable	Estimate	S. E.	Est./S.E.	p
Intercept	5.04	1.15	4.37	0.00

According to Table 1, the variance of students' gains in student level is 1247.87, while the variance of students' gains in school level is 30.99. According to these values, the intra-class correlation coefficient is calculated as 0.02. According to this value, it can be concluded that approximately 2% of the differences in the gain scores observed among the students arise from the difference in the average gain scores between the schools and 98% of it originates from differences in student level. According to Table 1, the average gain score of students is 5.04. However, there is a significant difference between schools in terms of the gain score ($p < .05$).

3.3. Findings on the effect of student and school characteristics on student gain scores

The results of the first level random intercept model are given in Table 2.

Table 2. The results of the first level random intercept model

Variable	Estimate	S. E.	Est./S.E.	p	Pratt index
Intercept	10.83	2.50	4.32	0.00	
Gender	-1.37	1.48	-0.92	0.36	-
Socioeconomic level	-0.13	0.04	-3.02	0.01	0.09
Variance components	Variance	df	Variance/df	p	
Residual	1241.31	43.38	28.61	0.00	
School level [u_{0j}]	26.03	9.91	2.63	0.01	

According to Table 2, while the socioeconomic level variable has a statistically significant effect on the gain score of students ($p < .05$), the gender variable does not have a statistically significant effect on the gain score ($p > .05$). According to the results, the effect of the socioeconomic level is negative and at a quite low level. According to this, a decrease of one standard deviation in the score in socioeconomic level resulted in an increase of 0.13 in the gain score of students. Besides, when the Pratt index value of the socioeconomic level is examined, it appears that this variable has no effect on the students' gain scores in a practical sense. The results of the regression model in which means are outcomes are given in Table 3.

Table 3. The results of the regression model in which means are outcomes

Variable	Estimate	S. E.	Est./S.E.	p
Intercept	7.09	2.90	2.44	0.01
School size	-0.00	0.01	-1.12	0.26
School resources	-0.09	0.47	-0.18	0.86
Variance components	Variance	df	Variance/df	p
Residual	1247.86	45.16	27.63	0.00
School level [u_{0j}]	29.87	10.95	2.73	0.01

According to Table 3, school size and resources do not have a statistically significant effect on students' gain score ($p > .05$). Therefore, the second level variables which are added to the model cannot explain the variance in the gain scores observed between schools.

3.4. Findings of the models using a categorical growth model

Frequencies related to the values calculated based on schools' growth categories are given in Appendix 2, and results based on statistics are given in Figure 3.

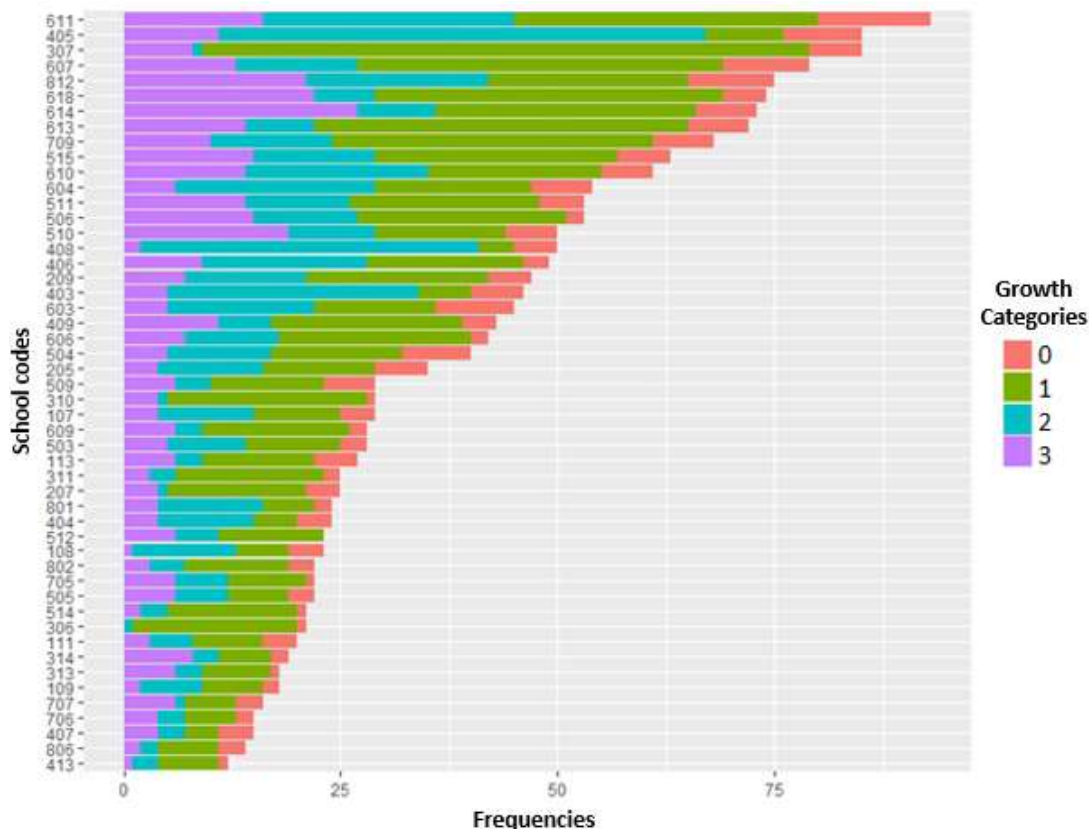


Figure 3. Frequency distribution of growth categories of schools

Upon examining schools in terms of growth categories according to Figure 3 and Appendix 2, it is seen that most students (43%) are in Category 1 and few (11.1%) are in Category 0. In other words, while most of the students are at the basic level, a small proportion of students consist of students who have downgraded their performance to a lower level. In addition, the school with the highest number of students (7%) who increased their performances is the school with the code of 614, and the lowest (0%) is the school with the code of 306. The school in which the highest number of students (5.9%) whose performance is downgraded to a lower level is the school with the code of 611 and the school with the lowest number (0%) is coded as 512.

3.5. Findings of differences between schools in terms of categorical growth model

The results regarding the models established to examine the differences between schools in terms of categorical growth model are given in Table 4. According to Table 4, the expected possibility of students to be in Category 2 than to be in Category 3 is $\frac{\exp\{0.17\}}{1 + \exp\{0.77\} + \exp\{0.17\} + \exp\{-0.53\}} = \frac{1.18}{4.92} = 0.23$; the expected possibility of being in Category 1 is $\frac{\exp\{0.77\}}{1 + \exp\{0.77\} + \exp\{0.17\} + \exp\{-0.53\}} = \frac{2.16}{4.92} = 0.44$; the expected possibility of being in Category 0 is $\frac{\exp\{-0.53\}}{1 + \exp\{0.77\} + \exp\{0.17\} + \exp\{-0.53\}} = \frac{0.58}{4.92} = 0.12$. The possibilities of all categories except for the Category 2 are statistically significant ($p < .05$). In addition, intercept variance is significant in the first and second categories, and there are significant differences among schools in these categories ($p < .05$).

Table 4. Results of random effects one-way ANOVA model

Fixed Effects	Estimate	S. E.	t	df	p
Category 0 for Intercept 1 $\beta_0(0)$					
Intercept 2 $\gamma_{00}(0)$	-0.53	.09	-5.67	49	<0.001
Category 1 for Intercept 1 $\beta_0(1)$					
Intercept 2 $\gamma_{00}(1)$	0.77	0.09	8.69	49	<0.001
Category 2 for Intercept 1 $\beta_0(2)$					
Intercept 2 $\gamma_{00}(2)$	0.17	0.13	1.31	49	0.20
Variance components	Standard deviation	Variance	df	X^2	p
Intercept 1 (0), $u_0(0)$	0.38	0.14	49	57.88	0.18
Intercept 1 (1) $u_0(1)$	0.43	0.19	49	92.87	<0.001
Intercept 1 (2) $u_0(2)$	0.77	0.60	49	160.16	<0.001

3.6. Findings of the effect of student and school characteristics on categorical growth

The results of the first level random intercept model are presented in Table 5. According to Table 5, female students are more than twice ($p < .05$) as likely to be in Category 1 instead of Category 3 ($\exp\{0.79\} = 2.20$) than male students. However, female students are 31% less likely ($p < .05$) to be in Category 2 rather than Category 3 compared to male students ($\exp\{-0.37\} = 0.69$). When the gender variable is controlled, an increase by one point in the socioeconomic level of the students is expected to increase the possibility of being in Categories 0, 1, or 2 instead of Category 3 by 0.01 points ($p < .05$). In this context, the data showed that the socioeconomic level has no effect on the students' categories in a practical sense. Another observation is that possibility of first level variables is not statistically significant in terms of other categories ($p > .05$). The results of the regression model in which means are outcomes are given in Table 6.

Table 5. The results of the first level random intercept model

Fixed Effects	Estimate	S. E.	t	df	p
Category 0 for Intercept 1 $\beta_0(0)$					
Intercept 2 $\gamma_{00}(0)$	-0.50	0.09	-5.48	49	<0.001
Gender $\beta_1(0)$, Intercept 2, $\gamma_{10}(0)$	0.23	0.16	1.36	1848	0.17
Socioeconomic level $\beta_2(0)$, Intercept 2, $\gamma_{20}(0)$	0.01	0.001	3.08	1848	.002
Category 1 for Intercept 1 $\beta_0(1)$					
Intercept 2 $\gamma_{00}(1)$	0.77	0.89	8.68	49	<0.001
Gender $\beta_1(1)$, Intercept 2, $\gamma_{10}(1)$	0.79	0.13	5.79	1848	<0.001
Socioeconomic level $\beta_2(1)$ Intercept 2, $\gamma_{20}(1)$	0.01	0.001	2.42	1848	.02
Category 2 for Intercept 1 $\beta_0(2)$					
Intercept 2 $\gamma_{00}(2)$	0.15	0.14	1.07	49	.29
Gender $\gamma_{20}(1)$ Intercept 2, $\gamma_{10}(2)$	-0.37	0.16	-2.35	1848	.02
Socioeconomic level $\beta_2(2)$, Intercept 2, $\gamma_{20}(2)$	0.01	0.003	3.25	1848	.001
Variance components	Standard deviation	Variance	df	X^2	p
Intercept 1 (0), $u_0(0)$	0.33	0.11	49	53.15	0.32
Intercept 1 (1) $u_0(1)$	0.44	0.19	49	91.14	<0.001
Intercept 1 (2) $u_0(2)$	0.79	0.62	49	157.63	<0.001

Table 6. The results of the regression model in which means are outcomes

Fixed Effects	Estimate	S. E.	t	df	p
Category 0 for Intercept 1 $\beta_0(0)$					
Intercept 2 $\gamma_{00}(0)$	-0.54	0.10	-5.43	47	<0.001
School size, $\gamma_{01}(0)$	0.0001	0.0001	1.59	47	0.12
School resources, $\gamma_{02}(0)$	-0.002	0.04	-0.04	47	0.96
Category 1 for Intercept 1 $\beta_0(1)$					
Intercept 2 $\gamma_{00}(1)$	0.76	0.09	8.62	47	<0.001
School size, $\gamma_{01}(1)$	-0.0001	0.0001	-0.15	47	0.88
School resources, $\gamma_{02}(1)$	-0.05	0.04	-1.01	47	0.32
Category 2 for Intercept 1 $\beta_0(2)$					
Intercept 2 $\gamma_{00}(2)$	0.14	0.12	1.12	47	0.27
School size, $\gamma_{01}(2)$	0.0001	0.0001	2.86	47	0.01
School resources, $\gamma_{02}(2)$	-0.09	0.05	-1.89	47	0.06
Variance components	Standard deviation	Variance	df	X^2	p
Intercept 1 (0), $u_0(0)$	0.39	0.15	47	56.99	0.15
Intercept 1 (1) $u_0(1)$	0.44	0.19	47	90.55	<0.001
Intercept 1 (2) $u_0(2)$	0.76	0.57	47	145.25	<0.001

According to Table 6, when the school resources are controlled for, an increase by one person in the size of schools would decrease the possibility of students' being in Category 2 rather than Category 3 by 0.09 ($p < .05$). Besides, the possibility of second level variables in terms of other categories is not statistically significant ($p > .05$).

4. DISCUSSION and CONCLUSION

In this study, the growth in students' comprehension skills was examined using the gain score and categorical growth models. Results showed that some students did not achieve sufficient gains to advance to higher performance levels. Although some schools' average gains were higher, their performance was still not significant enough in terms of tests' standards. Moreover, the analyses demonstrated that the student gain scores and growth categories varied significantly among the schools. In addition, the study was able to determine student and school characteristics that have an impact on the students' gain scores and categorical growth.

According to the results obtained from the gain score and categorical growth models, there was a significant difference between schools and students. When the average gain score values regarding schools and students were examined, the values proved to be positive. In this sense, it can be said that the students increased their scores from sixth to seventh grade in general. This is an expected situation upon considering the structure of comprehension skills in language (Crawford et al., 2001; Herbers et al., 2012). However, the results obtained from the gain model need to be evaluated cautiously (Betebenner & Linn, 2009; Castellano & Ho, 2013; Pike, 1991). The reason for this is to determine with the help of gain model how much the student scores have increased or decreased over one year. In other words, there is no information concerning the starting positions of the students in this model. In this case, the results of the model can be affected by the problem known as floor and ceiling effect (Rock and Pollack, 2002).

According to the results obtained from the categorical growth model, students were most likely to perform at the basic level in both the sixth and seventh grade. The lowest possibility was that students in the sixth grade performed at the proficient level, but in the seventh grade, they performed at the basic level. This can be considered as an expected result, given that the average performance of the participants in the study is low. In this sense, considering the fact that students' showing growth corresponds to improving in terms of levels in the categorical growth model (Ryser & Rambo-Hernandez, 2013), it can be stated that students within the context of the study have a low level of performance in achieving the proficient standards. Thus, although student growth has increased, this increase is not at the proficient level according the results of this study.

This study showed that gender had no effect on the significant difference in gain scores. The reason why it does not have any effect on the difference in the gain scores of students may be due to the ceiling effect; if the female students are more successful than the male students, it is expected that females' gain scores would be less. The indicators obtained from the categorical growth model revealed that female students are more likely to improve from the basic level to the proficient level than the male students. This situation between the female students and male students in terms of growth categories are in line with some studies in the literature (Denton & West, 2002; Kurdek & Sinclair, 2001). The fact that female students are more successful compared to male students also show similarity in terms of academic performance of students (Anıl, Özer – Özkan, & Demir, 2015; Büyüköztürk et al., 2014; Taş et al., 2016) in Turkey. This indicates that female students continue to display higher performance than expected at advancing grade levels than male students. This success may have resulted from the increase in the projects and programs especially intended for girls' education in recent years in Turkey.

It was determined that the socioeconomic level of the students had a significant negative effect on the gain scores, but this effect was not significant in a practical sense. At the same time, students' socioeconomic level was determined to have a significant negative effect on the possibility of passing towards from the basic level to the proficient level. This can be an example of the state of academically resilient students. Turkish researchers (e.g. Dinçer & Oral, 2010; Yavuz & Kutlu, 2016) studying academically resilient students have found that evidence to suggest that students can be academically successful in spite of the disadvantages socioeconomic level cause. Considering the participants of this study, it can be stated that the students, in general, came from a socioeconomically disadvantaged district.

The findings of this study related to school characteristics are in parallel with some studies in the literature (Glewwe, Hanushek, Humpage, & Ravina, 2011; Leithwood & Jantzi, 2009). In this context, it can be stated that the physical characteristics of schools do not play an important role in students' gain scores. It can be expressed that a similar situation can be observed in the academic achievements of students in Turkey. In addition, the fact that the study group was gathered from only one district of Ankara and that the district schools' have similar physical resources may have affected the results. It has been shown that school size and school resources which are among school characteristics, displayed no significant difference in growth categories except for the category of "remaining at the proficient level". According to the results of the analysis, the decrease in the size of the school increases the possibility of the students to remain at the proficient level. This is in line with the results of the meta-analysis study conducted by Leithwood and Jantzi (2009) on the relationship between school size and student achievement, but not in parallel with Stevenson (1996)'s study. Furthermore, given that school size has no effect on other categorical growths in this study, it can be stated that this finding is not effective in a practical sense.

It is suggested in this study that there are differences between the students and the schools in terms of their growth level. Given the different aspects gained about students' performance with

these models, it is recommended to utilize different growth models in schools. In these assessments, especially because there are performance levels that have a full meaning/equivalence for the education shareholders rather than the scores of the students, the outputs of these models can be helpful for the education shareholders (Slaughter, 2008). Thus, educational stakeholders can monitor the student growth and determine whether or not the students are at the expected level of performance; thereby relevant educational institutions can take effective measures. Considering the results of this study concerning the characteristics of students and schools, for the purpose of increasing student growth, educational stakeholders may construct policies for other school characteristics that can assist in student growth.

The results of this study should be evaluated within the scope of its limitations. Since vertically scaled tests could not be used in the study and pseudo common items are used, it would be more appropriate for future researchers to use vertically scaled tests in conducting studies where students are monitored. It would be also efficient to use tests including more items and giving more information with a larger theta range. In addition, because the study is conducted in Altındağ district of Ankara, which has a low socioeconomic status and homogenous features within the district, future researchers may want to conduct a similar study in a more heterogeneous and larger study group. Aside from this, the grade level may also be included in the analysis. In future studies to be conducted, a similar study could investigate for reading comprehension, which is more general instead of language comprehension skills. In addition, a limited number of variables regarding the student and school characteristics were used in this study. For this reason, a similar study could be carried out by including other student and school characteristics that could be predictive of student growth.

Acknowledgements

This paper was extracted from the first author's Ph.D. thesis

ORCID

Hatice Cigdem YAVUZ  <https://orcid.org/0000-0003-2585-3686>

Ömer KUTLU  <https://orcid.org/0000-0003-4364-5629>

5. REFERENCES

- Anıl, D., Özer Özkan, Y. ve Demir, E. (2015). *PISA 2012 araştırması ulusal nihai rapor*. T.C. Millî Eğitim Bakanlığı, Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara.
- Arnold, D. H., & Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annual Review of Psychology*, 54, 517–545.
- Ayral, M., Özdemir, N. ve Sadıç, Ş. (2011). *Altındağ İlçesi Öğrenme Düzeyi Araştırması Raporu*. Altındağ Rehberlik ve Araştırma Merkezi, Ankara. Retrieved February 20, 2017 from http://altindagram.meb.k12.tr/meb_iys_dosyalar/06/01/334395/dosyalar/2012_12/18012758_renmedzeyiartrmas_1.pdf
- Betebner, D. W. (2009). *Growth, standards and accountability*. Retrieved February 25, 2017 from the National Center for the Improvement of Educational Assessment: http://www.nciea.org/publications/normative_criterion_growth_DB08.pdf
- Briggs, D., & Betebner, D. W. (2009, April). *Is growth in student achievement scale dependent?* Paper presented at the invited symposium Measuring and Evaluating. Changes in Student Achievement: A Conversation About Technical and Conceptual Issues at the annual meeting of the National Council for Measurement in Education, San Diego, CA.
- Browne, W. J. & Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1, 473–514.

- Bursal, M., Buldur, S. & Dede, Y. (2015). Alt sosyo-ekonomik düzeyli ilköğretim öğrencilerinin 4-8. sınıflar fen ve matematik ders başarıları: Cinsiyet perspektifi [Science and mathematics course success of elementary students in low socio-economic status among 4th-8th grades: Gender perspective]. *Eğitim ve Bilim*, 40(179) 133-145.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş. ve Atar, H. Y. (2014). *TIMSS 2011 ulusal matematik ve fen raporu- 8. Sınıflar*. T.C. Millî Eğitim Bakanlığı, Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara.
- Castellano, K. E., & Ho, A. D. (2013). *A practitioner's guide to growth models*. Washington, DC: Council of Chief State School Officers. Retrieved February 20, 2017 from: http://scholar.harvard.edu/files/andrewho/files/a_practitioners_guide_to_growth_models.pdf
- Chavent, M., Kuentz-Simonet, V., Labenne, A., & Saracco, J. (2014). *Multivariate analysis of mixed data: The PCAmixdata R package*. arXiv preprint arXiv:1411.4911.
- Cheti, N., & Birgitta, R. (2012). *The effect of school resources on test scores in England*, ISER Working Paper Series, No. 2012-13.
- Crawford, L., Tindal, G., & Steiber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, 7(4), 303–323.
- Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade* (NCES 2002-125). Washington, DC: National Center for Education Statistics.
- Dinçer, M. A. ve Oral, I. (2010). *Türkiye'de devlet liselerinde akademik yılmazlık profili: PISA 2009 Türkiye verisinin analizi*. İstanbul: Eğitim Reformu Girişimi.
- Ergin-Aydemir, S. ve Sünbül, Ö. (2016). Matematik bilişsel gelişiminin örtük büyüme modeli ile izlenmesi [Monitoring the mathematics cognitive development with the latent growth modeling]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 16(1), 20-40.
- Erman-Aslanoğlu, A. ve Kutlu, Ö. (2015). Factors related to the reading comprehension skills of 4th grade students according to data of PIRLS 2001 Turkey. *Journal of Educational Sciences Research*, 5(2), 1-18.
- Glewwe, P. W., Hanushek, E. A., Humpage, S. D., & Ravina, R. (2011). *School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010*. Working Paper 17554. Cambridge, MA: National Bureau of Economic Research.
- Goldschmidt, P., Choi, K., & Beaudoin, J. P. (2012). *Growth model comparison study: Practical implications of alternative models for evaluating school performance*. Washington, DC: Council of Chief State School Officers.
- Gong, B. (2004). *Models for using student growth measures in school accountability*. Paper presented at the Council of Chief State School Officers' "Brain Trust" on Value-added Models, Washington, DC.
- Gonzalez, J. & Wilberg, M. (2017). *Applying Test Equating Methods Using R*. New York: Springer International Publishing.
- Güzle-Kayır, Ç. ve Erdoğan, M. (2015). The variation in Turkish students' reading skills based on PISA 2009: The effects of socio-economic and classroom-related factors. *International Online Journal of Educational Sciences*, 7(4), 80 – 96.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for the item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26, 3–24.
- Hanushek, E. A. (2006). School resources. in Hanushek, E. A., and Welch, F. (eds) *Handbook of the Economics of Education*, (Amsterdam: Elsevier) 865-908
- Heck, R. H. (2006). Assessing school achievement progress: Comparing alternative approaches. *Educational Administration Quarterly*, 42(5), 667-699.

- Herbers, J. E., Cutuli, J. J., Supkoff, L. M., Heistad, D., Chan, C.-K., Hinz, E., & Masten, A. S. (2012). Early reading skills and academic achievement trajectories of students facing poverty, homelessness, and high residential mobility. *Educational Researcher*, 41, 366–374.
- Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology*, 100(1), 1–14.
- Husain, M., & Millimet, D. (2009). The mythical “boy crisis”? *Economics of Education Review*, 28, 38-48.
- Krueger, A. B., & Lindahl (2001). Education for growth: Why and for whom?, *Journal of Economic Literature*, 39(4), 1101–1136.
- Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology*, 93(3), 451–455.
- Kutlu, Ö., Yıldırım, Ö., Bilican, S. ve Kumandaş, H. (2011). İlköğretim 5. sınıf öğrencilerinin okuduğunu anlamada başarılı olup olmama durumlarının kestirilmesinde etkili olan değişkenlerin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(1), 132-139.
- Laird, E. (2008). *Tapping into the power of longitudinal data: A guide for school leaders*. Retrieved January 21, 2018 from <http://www.dataqualitycampaign.org/>.
- Leithwood, K., Edge, K., & Jantzi, D. (1999). *Educational accountability: The state of the art*. Gütersloh, Germany: Bertelsmann Foundation.
- Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). *Standard setting: A Bookmark approach*. In D. R. Green (Chair), IRT-based standard setting procedures utilizing behavioral anchoring. Symposium conducted at the Council of Chief State School Officers National Conference on Large-Scale Assessment, Phoenix AZ.
- Liu, Y., Zumbo, B. D., & Wu, A. D. (2014). Relative importance of predictors in multilevel modeling. *Journal of Modern Applied Statistical Methods*, 13(1), 1-22.
- Luppescu, S. (2005). Virtual equating. *Rasch Measurement Transactions*, 19(3), 10-25.
- McCoach, B. D., O’Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers: A hierarchical linear model of children’s reading growth during the first 2 years of school. *Journal of Educational Psychology*, 98, 14-28.
- Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children’s growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities*, 44, 472- 488.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2001). PIRLS 2001 International Report. Chestnut Hill, MA: International Study Center. Retrieved February 20, 2017 from https://timssandpirls.bc.edu/pirls2001i/pdf/p1_IR_book.pdf.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus User’s Guide* (Eighth Edition). Los Angeles, CA: Muthén & Muthén
- National Assessment Agency. (2008). *14-19 Reforms*. Retrieved February 3, 2017 from: <http://www.naa.org.uk/>
- Nayır, F. (2013). Eğitimde kalite geliştirme sürecinde okul değerlendirmenin rolü [The role of school evaluation in the process of improving quality of education]. *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 11(2), 119-134.
- Nese, J., Biancarosa, G., Anderson, D., Lai, C., Alonzo, J., & Tindal, G. (2012). Within- year oral reading fluency with CBM: A comparison of models. *Reading & Writing*, 25(4), 887-915.
- No Child Left Behind Act of 2001, P. L. 107-110, 20 U.S.C. 6319 (2002).

- OECD (2014a). *PISA 2012 Results: What Students Know and Can Do – Student Performance in Mathematics, Reading and Science* (Volume I, Revised edition, February 2014), PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264201118-en>
- OECD. (2014b). PISA 2012 technical report. Paris: OECD Publications.
- OECD (2016). PISA 2015 Results (Volume I): Excellence and Equity in Education, PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264266490-en>
- Osborne, J. W. (2000). Advantages of hierarchical linear modeling. *Practical Assessment, Research, and Evaluation*, 7(1), 1-3.
- Özer Özkan, Y. (2016). Okulları başarılarına göre sınıflandırmada etkili olan değişkenlerin PISA 2012 Türkiye verileri aracılığıyla incelenmesi [The impact of school properties to mathematics literacy in the PISA 2012 Turkey sample]. *International Online Journal of Educational Sciences*, 8(2), 117-130.
- Palardy, G. J. (2008). Differential school effects among low, middle, and high social class composition schools: A multiple group, multilevel latent growth curve analysis. *School Effectiveness and School Improvement*, 19, 21–49.
- Pike, G. R. (1991). Using structural equation models with latent variables to study student growth and development. *Research in Higher Education*, 32(5), 499-524.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publision.
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2011). *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International.
- Rock, D. A., & Pollack, J. M. (2002). *A model-based approach to measuring cognitive growth in pre-reading and reading skills during the kindergarten year*. ETS Research Report Series, 2002(2).
- Ryser, G., & Rambo-Hernandez, K. (2014). Using growth models to measure school performance: Implications for gifted learners. *Gifted Child Today*, 37(1), 17–23.
- Shin, T., Davison, M. L., Long, J. D., Chan, C., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning and Individual Differences*, 23, 92-100.
- Skibbe, L. E., Connor, C. M., Morrison, F. J., & Jewkes, A. M. (2010). Schooling effects on preschoolers' self-regulation, early literacy, and language growth. *Early Childhood Research Quarterly*, 26, 42-49.
- Slaughter, R. (2008). *Measuring middle school achievement growth with student growth percentile methodology*. NERA Conference http://digitalcommons.uconn.edu/nera_2008/20
- Stevenson, H. (2006) Moving towards, into and through principalship: Developing a framework for researching the career trajectories of school leaders. *Journal of Educational Administration*, 44(4), 408-420.
- Taş, U. E., Arıcı, Ö., Özarkan, H. B. ve Özgürlük, B. (2016). *PISA 2015 ulusal raporu*. T.C. Millî Eğitim Bakanlığı, Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara.
- Türkoğlu, M. E. ve Aypay, A. (2015). Özel okul öğretmenlerinin öğretmen hesap verebilirliğine dair düşünceleri [Private school teachers' thoughts about teacher accountability]. *Eğitimde Politika Analizi Dergisi*, 4(1), 7-32.
- U.S. Department of Education. (2010, June 24). Race to the Top Assessment Program. Announcement of Race to the Top Program by U.S. Department of Education. Retrieved February 18, 2017 from <http://www2.ed.gov/programs/racetothetop-assessment/index.html>

- Welch, C., Dunbar, S., & Rickels, H. (2016). *Measuring student growth in Iowa with Iowa assessments*. Retrieved February 25, 2017 from <https://itp.education.uiowa.edu/ia/documents/Measuring-Student-Growth-in-Iowa-with-the-Iowa-Assessments.pdf>
- Yapar, T. (2014). *İngilizce okuma becerisindeki gelişimin Madde Tepki Kuramı ve örtük büyüme modellemesiyle incelenmesi* [An investigation of the development in English reading skill by using item response theory and latent growth modeling]. (Unpublished doctorate thesis). Hacettepe Üniversitesi, Ankara.
- Yavuz, H. Ç., & Kutlu, Ö. (2016). Ekonomik Bakımdan Dezavantajlı Öğrencilerin Akademik Yılmazlık Düzeylerinin Bazı Koruyucu Faktörler Açısından İncelenmesi [Investigation of the factors affecting the academic resilience of economically disadvantaged high school students]. *Eğitim ve Bilim*, 41(186), 1-19.

6. APPENDICES

Appendix 1. Descriptive statistics related to the values calculated according to the gain score model of schools

School code	N	Mean	Standard Deviation	Min	Max	School code	N	Mean	Standard Deviation	Min	Max
107	29	-7.02	25.92	-57.87	40.36	705	22	2.90	31.66	-43.04	68.08
108	23	-1.15	32.03	-54.53	74.77	706	15	4.02	45.16	-54.53	84.08
109	18	6.71	35.72	-40.26	84.08	707	16	18.83	56.72	-55.35	112.18
111	20	6.15	30.01	-40.73	47.37	709	68	7.51	32.79	-64.82	83.36
113	27	3.67	33.32	-64.82	70.66	801	24	-4.98	37.64	-53.75	97.83
205	35	0.49	31.59	-55.35	67.54	802	22	6.37	39.25	-61.83	70.66
207	25	8.89	31.09	-55.35	82.10	806	14	2.65	39.59	-61.83	57.33
209	47	-0.71	32.60	-53.09	81.89	812	75	10.07	40.85	-55.35	112.18
306	21	0.29	28.38	-40.73	47.37						
307	85	0.10	32.22	-57.87	84.08						
310	29	10.39	28.33	-42.15	74.77						
311	25	17.84	33.36	-54.02	82.10						
313	18	0.09	28.78	-54.53	54.93						
314	19	18.18	38.21	-53.75	84.08						
403	46	-8.16	34.13	-59.75	67.54						
404	24	-0.97	30.17	-43.04	67.54						
405	85	-0.53	31.36	-54.02	73.42						
406	49	4.37	35.01	-56.46	128.36						
407	15	-2.51	45.49	-54.53	97.83						
408	50	-12.95	29.00	-59.75	68.35						
409	43	8.49	34.22	-55.35	70.66						
413	12	6.83	30.57	-27.53	81.89						
503	28	3.14	38.85	-57.87	96.45						
504	40	2.09	38.19	-55.35	101.52						
505	22	15.38	39.73	-54.02	70.66						
506	53	14.72	37.23	-57.87	112.18						
509	29	-0.64	39.63	-53.75	88.10						
510	50	15.32	37.12	-54.53	96.45						
511	53	17.45	44.22	-64.82	101.52						
512	23	15.42	39.86	-48.02	82.10						
514	21	-3.06	31.10	-55.35	70.66						
515	63	5.69	38.19	-57.87	84.08						
603	45	-6.01	36.61	-57.87	97.83						
604	54	-3.98	33.91	-57.87	84.08						
606	42	0.83	31.75	-64.82	61.28						
607	79	8.20	31.36	-57.87	84.08						
609	28	8.34	36.31	-55.35	112.18						
610	61	4.01	34.11	-61.83	144.89						
611	93	4.98	37.60	-59.75	101.52						
613	72	4.59	31.89	-61.83	74.77						
614	73	16.63	37.73	-48.02	112.18						
618	74	20.27	37.77	-44.51	127.46						
Total	2004	5.24	35.78	-64.82	144.89						

Appendix 2. Frequencies related to the values calculated based on schools' growth categories

School code	N	f (0)	%	f (1)	%	f (2)	%	f (3)	%	School code	N	f (0)	%	f (1)	%	f (2)	%	f (3)	%
107	29	4	1.8	10	1.2	11	2.1	4	1.0	611	93	13	5.9	35	4.1	29	5.4	16	4.1
108	23	4	1.8	6	.7	12	2.2	1	.3	613	72	7	3.2	43	5.0	8	1.5	14	3.6
109	18	2	.9	7	.8	7	1.3	2	.5	614	73	7	3.2	30	3.5	9	1.7	27	7.0
111	20	4	1.8	8	.9	5	.9	3	.8	618	74	5	2.3	40	4.6	7	1.3	22	5.7
113	27	5	2.3	13	1.5	3	.6	6	1.6	705	22	1	.5	9	1.0	6	1.1	6	1.6
205	35	6	2.7	13	1.5	12	2.2	4	1.0	706	15	2	.9	6	.7	3	.6	4	1.0
207	25	4	1.8	16	1.9	1	.2	4	1.0	707	16	3	1.4	6	.7	1	.2	6	1.6
209	47	5	2.3	21	2.4	14	2.6	7	1.8	709	68	7	3.2	37	4.3	14	2.6	10	2.6
306	21	1	.5	19	2.2	1	.2	0	0.0	801	24	2	.9	6	.7	12	2.2	4	1.0
307	85	6	2.7	70	8.1	1	.2	8	2.1	802	22	3	1.4	12	1.4	4	.7	3	.8
310	29	1	.5	23	2.7	1	.2	4	1.0	806	14	3	1.4	7	.8	2	.4	2	.5
311	25	2	.9	17	2.0	3	.6	3	.8	812	75	10	4.5	23	2.7	21	3.9	21	5.4
313	18	1	.5	8	.9	3	.6	6	1.6										
314	19	2	.9	6	.7	3	.6	8	2.1										
403	46	6	2.7	6	.7	29	5.4	5	1.3										
404	24	4	1.8	5	.6	11	2.1	4	1.0										
405	85	9	4.1	9	1.0	56	10.5	11	2.8										
406	49	3	1.4	18	2.1	19	3.6	9	2.3										
407	15	4	1.8	4	.5	3	.6	4	1.0										
408	50	5	2.3	4	.5	39	7.3	2	.5										
409	43	4	1.8	22	2.6	6	1.1	11	2.8										
413	12	1	.5	7	.8	3	.6	1	.3										
503	28	3	1.4	11	1.3	9	1.7	5	1.3										
504	40	8	3.6	15	1.7	12	2.2	5	1.3										
505	22	3	1.4	7	.8	6	1.1	6	1.6										
506	53	2	.9	24	2.8	12	2.2	15	3.9										
509	29	6	2.7	13	1.5	4	.7	6	1.6										
510	50	6	2.7	15	1.7	10	1.9	19	4.9										
511	53	5	2.3	22	2.6	12	2.2	14	3.6										
512	23	0	0.0	12	1.4	5	.9	6	1.6										
514	21	1	.5	15	1.7	3	.6	2	.5										
515	63	6	2.7	28	3.3	14	2.6	15	3.9										
603	45	9	4.1	14	1.6	17	3.2	5	1.3										
604	54	7	3.2	18	2.1	23	4.3	6	1.6										
606	42	2	.9	22	2.6	11	2.1	7	1.8										
607	79	10	4.5	42	4.9	14	2.6	13	3.4										
609	28	2	.9	17	2.0	3	.6	6	1.6										
610	61	6	2.7	20	2.3	21	3.9	14	3.6										
Total	2004	222	100	861	100	535	100	386	100										