Education and Science

Vol 45 (2020) No 202 79-91

The Implementation of the Value-Added Assessment to Determine School Effectiveness *

Sedat Şen [1], İbrahim Yıldırım [2], Mehmet Fatih Karacabey [3]

Abstract

The aim of this study was to introduce the value-added assessment model, one of the approaches used in school and teacher evaluation, and to demonstrate its implementation on a real data set. In the study, success ratings for the schools were assessed through the value-added assessment model using TEOG (transition from primary to secondary education) exam scores of 539 secondary schools in Şanlıurfa province of Turkey. The school ratings based on the exam averages were also included. The differences were found between the ratings based on the value-added scores obtained in the study and the traditional ones depending only on the exam averages. It was concluded that it was not accurate to evaluate the school success solely by a specific exam average, and the value-added assessment approach, revealing how much the schools were differentiated from one exam to another, was decided to offer more objective results. The present study, serving as a model for the use of value-added assessment approach in Turkey, offers suggestions to assist school and teacher evaluation to be planned in the future.

Keywords

School effectiveness
Value-added assessment
Value-added assessment models

Article Info

Received: 07.05.2018
Accepted: 02.13.2020
Online Published: 04.22.2020

DOI: 10.15390/EB.2020.8023

Introduction

Education has been carried out in schools worldwide despite some radical views regarding the training of individuals and educational environments. According to the report released by UNESCO in 2019, the proportion of students enrolling in primary education worldwide is 92% while that of the graduates is 84% (UNESCO, 2019). In Turkey, the schooling rates of 89.77% in primary education, 56.63% in secondary education and 18.85% in higher education in 2005-2006 academic year have increased to 91.92% in primary education, 93.28% in secondary education and 44.10% in higher education in 2018-2019 academic year (Ministry of National Education [MoNE], 2019). While the schools are financed by the state in numerous countries, private enterprise is dominant in some others. Regardless of being supported by the government or private financial resources, it can be defined as a problem situation to determine whether the financial resources allocated to schools yield the expected returns. The clarification of this problem situation will be an important step towards revealing the meaning and function of the school in terms of effectiveness.

* This article is orally presented at the 16th International Conference on Teaching, Education & Learning.

[1] Harran University, Faculty of Education, Department of Educational Sciences, Turkey, sedatsen06@gmail.com
[2] Gaziantep University, Gaziantep Faculty of Education, Department of Educational Sciences, Turkey, iyildirim84@gmail.com
[3] Harran University, Faculty of Education, Department of Educational Sciences, Turkey, karacabeyakademi@gmail.com

Although it has been widely used, the concept of effectiveness gains different meanings according to the environment in which the organization is located, the purposes it aims to achieve, and its stakeholders and varies depending on these variables. Effectiveness is the degree of achieving and realizing the goals of the organization as a result of its activities, namely a sort of performance indicator regarding the extent of achieving its goals (Yükçü & Atağan, 2009).

Barnard (1938) emphasizes the importance of formal organization in terms of organizational effectiveness, and claims that it is possible to improve the effectiveness and efficiency of an organization through formal structuring. He highlights the importance of balancing within the organization, adapting to the external environment and analysing the effects of managerial performance as a unitary system of the organization for the continuity of organization's existence. According to him, an organization requires a formal structure as its stakeholders need to work for the continuity, effectiveness and efficiency of the organization. In this regard, managers can change the actions and goals of employees through influence and control mechanisms, and develop practices and policies to protect the organization's internal balance against the environmental forces (Barnard, 1938).

The effectiveness of an organization is usually dealt with two approaches, namely a goal-oriented approach (the success of organization in achieving its goals) or a natural system approach (the survival of organization) (Rowan, 1985). When the effectiveness of schools, which are educational organizations, is analysed based on the goal-oriented approach, whether or how much they have reached the pre-determined learning outcomes is an indicator for their effectiveness. According to the natural system approach that investigates the effectiveness from a wider framework, the effectiveness of a school is scrutinized in a way to cover the social and life skills that have been developed as a result of the training received by the individual together with their values and behaviours (Karip & Köksal, 1996). The utmost problem encountered in evaluating the effectiveness of schools is that it is very difficult to determine the extent of goal achievement as the educational systems has short and long term objectives, latent and apparent goals. However, no matter how difficult it may be, it would not be incorrect to determine how much the pre-determined goals and objectives have been achieved in measuring school effectiveness (Creemers & Kyriakides, 2008). As a result, school effectiveness is the degree to reach its pre-determined goals, and therefore, an effective school is the one with the characteristics that have a positive effect on achieving its goals in terms of institutional, physical and human resources and so on (Botha, 2010). In short, it can be claimed that school effectiveness is related to how much student learning takes place.

Many initiatives have emerged around the world in determining school effectiveness. In addition to evaluating school effectiveness through the grade point averages, the assessments considering the teaching competence levels of teachers, educational inputs of the school, the type/amount of materials through data envelopment analysis and similar methods were also suggested. With a similar approach to data envelopment analysis, EARGED (2010) indicated that the main dimensions of "School Management", "Competence of Employees", "Quality of Education and Training Processes" and "Support Services" should be taken into consideration in evaluating school success. Adams (1993) noted that while the need for quality educational institutions was focused on during the 80s and 90s, it was not possible to meet on a common ground as every stakeholder defined the quality differently, Hesapçıoğlu (2006) pointed out that the concept of quality is intertwined with the notions of effectiveness, efficiency, performance and competence. Therefore, it is possible to assert that the answer to the question of what the most significant indicator of school effectiveness is student achievement to a great extent. Therefore, it is more plausible to focus directly on student achievement instead of dealing with so many different variables and uncontrollable cases. As a result, student achievement draws attention to be one of the most prominent elements of quality indicators for school quality and learning process outcomes.

For educators, it has been an object of interest for a long time to develop school performance criteria to improve the quality of education by considering student achievement. Based on the existing practice, MoNE determines the success of each school according to the scores of students in central government examinations (TEOG and YGS) aiming at transition from primary to secondary education and transition to higher education respectively. The evaluation of school success according to the result of a single exam (usually at the end of the year) has been criticized by both researchers and practitioners (Beardsley, 2008). Another practice frequently mentioned by school administrators is to compare the average scores of the present and previous years to decide whether the school is more successful. Considering the number of out of school factors affecting student achievement, it can be alleged that it is improper to evaluate school success through the average scores. In addition, these practices only address student achievement at the end of academic year by ignoring the one at the very beginning. Controlling the differences and determining how much each student has progressed from the very beginning of the semester to the end will be more accurate to determine school success as the students may be at different levels at the outset due to their learning experiences, socio-economic status and individual differences. In recent years, a great many methods have been developed based on the amount (improvement) of student achievement in evaluating school and/or teacher success. Value-added assessment systems are at the top of these ways. The value-added assessment (the VAA) approach is an objective assessment system that allows to attribute the level of improvement to schools or teachers by controlling the differences between the students and schools and estimating how much each student has improved from one exam to another.

Value-Added Assessment

There has been a growing interest in school effectiveness and accountability in the late 20th and the early 21st centuries though school effectiveness research dated back to the very beginning of the twentieth century. Especially with the No Child Left Behind Act, which was introduced in the USA in 2001, it has been aimed to measure the academic achievement of the students through the adoption of state-wide tests which has been implemented since the early 1900s (Kelly, 1916) by all the others (Beardsley, 2008). The first applications of this state-wide assessment focused on the present condition of students. The current status approach compare students at different grade levels at a single point in time (Doran & Izumi, 2004). The educators agree that the test scores obtained at single point is not a useful way to estimate school effects on student performance. The current status methods do not take into account the socio-economic factors when evaluating school effectiveness. Although these methods are at the heart of the school accountability system, there are many reasons why these methods are improper and cannot be used for school comparisons. First of all, the students enrol in school with different backgrounds and they are not distributed to schools randomly (Doran & Izumi, 2004). Therefore, it creates injustice for the comparison of schools that are advantageous and disadvantageous in terms of socio-economic status. Another rationale is that current status methods depend on increasing achievement through supplementing. In other words, the current achievement of students is based not only on their schools but also on their learning from their previous schools. Indeed, Kerbow (1996) affirmed that the students who change their schools affect the system and success of their new school. In addition, Temple and Reynolds (1999) voiced that the achievement of students who change their schools has decreased. Therefore, the current school of education cannot be held responsible for student achievement or failure. It is frequently emphasized in the literature that the accountability systems based on the current status are erroneous and induce faulty decisions in determining school quality (Drury & Doran, 2003). According to Drury and Doran (2003), there are a number of reasons behind the invalidity and delusiveness of current status method. These are involving the combined effect of family, community, background and previous schools, involving the cumulative student achievement of previous schools, and inadequacy in reflecting achievement at a specific grade and level. The shortcomings of this method have led to the emergence of an alternative way of evaluating school

effectiveness in the accountability system. This new method focuses on the "improvement" of the students throughout the year at school. Doran and Izumi (2004) pointed out that it is more logical to measure the progress of students over time from one assessment to the next in terms of "learning", which means "change" instead of showing how the level of knowledge of different groups of students is.

Researchers have developed value-added analysis methods that allow using the individual achievement of students longitudinally in determining effective schools. According to the definition of Tekwe et al. (2004, p.12), "value-added is a term used to label methods of assessment of school/teacher performance that measure the knowledge gained by individual students from one year to the next and then use that measure as the basis for a performance assessment system." The developers of the value-added analysis (the VAA) method suggest that it is fairer than the current status method in terms of considering the differentiation of achievement rather than focusing on a single point of achievement (Hanushek, 1972; Sanders, Saxton, & Horn, 1997; Tekwe et al., 2004). The main purpose of the VAA methods is to decide on school effectiveness based on the improvement of the students. Student improvement is determined by the gain scores based on the difference of the students' test results for consecutive years. The gain scores for each student are used in value-added assessment (the VAA) systems. In short, the VAA system is based on two basic ideas: (1) every student is nested in schools and (2) the performance change from one year to the other. Another advantage of the VAA system over the current status approach is that it can control the effect of school and student variables affecting the test scores. In this way, the effects of sociocultural characteristics and the previous experiences of the students can be reduced.

Value-Added Assessment Models

Meyer and Dokumaci (2010) indicated that the capability of the VAA system to produce high-quality value-added scores depends on the quality and suitability of the output to measure student achievement, the presence of longitudinal data and the design of the value-added assessment model (VAM). The VAMs usually represent the statistical models used to investigate the effect of educational parameters such as teachers and schools on student achievement. These models enable examining the improvement of students over time as a function of school and student characteristics. The VAMs have recently attracted much attention of the researchers and policy makers of education. Some of them based on the gain scores have been developed with hierarchical models with random effects. Hierarchical models used for the VAA analysis assume that the schools are random unlike the current status approach, which includes regression methods and presumes that schools distribute normally. As in the current status based analysis, statistical control can be used in hierarchical models to prevent confounding variables on the relationship between student achievement and school effectiveness (Sanders, 2000).

Hanushek (1972) is the first researcher to recommend using the VAMs in school evaluation systems. Sanders et al. (1997), on the other hand, were the first researchers who could apply the VAA approach on a state basis by creating the Tennessee VAA system (TVAAS). According to the extensive VAA study by McCaffrey, Lockwood, Koretz, and Hamilton (2003), while the first VAMs presumed fixed effects (Hanushek, 1972; Murnane, 1975), the later ones (hierarchical models and TVAAS model) assumed random effects. The main VAMs include simple fixed effects model, hierarchical linear models, and layered mixed effects model. McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004) concluded that a prevailing model that they called "multivariate model" has usually been followed which directly model the full common distribution of all student outcomes in analyzing the effects of teachers or schools through longitudinal data in the VAA approach, and there are also alternative model approaches that enable such data to be analyzed in the value-added assessment system. The alternative approaches include covariate adjustment models based on the regression of current scores with the previous ones, gain scores model that address year-end gains as the outcomes and cross-classified models. Many of these models require equated and standardized test scores in multiple numbers for the same student (Ballou, Sanders, & Wright, 2004; Doran & Cohen, 2005). On the other hand, the VAMs

that do not require an equation assumption have taken place in the literature in recent years (Mariano, McCaffrey, & Lockwood, 2010). The fact that consecutive exams used in educational institutions are mostly not equated or standardized on the same scale prevents the implementation of the VAMs that require an equation assumption between the exams. The analysis of non-equated exams through these models leads to biased and erroneous results (Briggs & Domingue, 2013; Briggs & Weeks, 2009). Generalized Persistence [GP] model which eliminates the limitations of the aforementioned models and does not require an equation assumption was developed by Mariano et al. (2010).

The Purpose and Significance of the Study

The "Teacher Performance Evaluation" implementation was carried out by Ministry of National Education in Turkey from the 23rd to the 31st of October, 2017 which had piloted in 132 schools in 12 provinces, but the country-wide authorized union decided counter action against the practice. In other words, this practice of MoNE within the scope of teacher evaluation was rendered inoperable by the counter action decision of the authorized union. The justification for the decision of the authorized union was based on teacher performance assessment including the evaluations of students, administrators, parents and the scores in an exam that teachers would take. Therefore, the proposed framework in this study, which has been implemented in a number of countries around the world, not only examines the feasibility of a teacher or school evaluation model in Turkey (particularly in Şanlıurfa), but also tries to implement a model in which teacher or school can be evaluated according to the improvement of student achievement.

The examination of research on school effectiveness in Turkey implies that it has mostly been attempted to be determined through data collection tools such as scales, questionnaires and interviews (Turhan, Şener, & Gündüzalp, 2017). The assessment of school effectiveness through the perceptions of school stakeholders such as administrators, teachers, students or parents using those instruments leads to subjectivity in evaluation. Therefore, there is a need to use more objective criteria. As a matter of fact, the factors that necessitate the present study are the fact that there is no objective and scientific method to hold the schools responsible for student achievement in the existing educational system, the value-added assessment (the VAA) approach offers a great many advantages and this approach has never been undertaken by educational policy makers in Turkey. This study, which was conducted depending on the aforementioned requirements, is considered to be crucial in terms of monitoring the learning outcomes of students in Şanlıurfa and using them in an objective school evaluation.

In this context, the aim of the study was to design the value-added assessment model, which has been used effectively in developed countries, in a form that will help MoNE to evaluate its affiliated educational institutions, and to test the prototype of this model on the data of Şanlıurfa. Another aim of the study was to contribute to develop educational policies and/or intervention strategies and to establish a basis for improving the quality of education. This study was aimed to introduce and apply the value-added assessment system providing valid, reliable and useful information for the evaluation of school achievement. In this regard, answers to the following questions were sought:

1. How are the secondary schools in Şanlıurfa province ranked in terms of effectiveness?
2. Do the single-exam success ratings and value-added score ratings of secondary schools in Şanlıurfa overlap?
3. Is the value-added assessment model functional based on the scores obtained from the secondary schools in Şanlıurfa?

The research findings will cover the results of the preliminary implementation of value-added assessment in the context of Şanlıurfa and will provide information about the feasibility of this approach.

Method

At least two estimates for each student are required to determine school effectiveness with the VAA analyses. In this study, the value-added scores of the included schools were calculated through the VAA using two standard test data obtained in one academic year. In this regard, the data of November and April TEOG exams (the Exam for the Transition from Basic Education to Secondary Education) in the 2015-2016 academic year were used. The schools were ranked according to the value-added scores obtained from the analyses. The total scores of the students who took the TEOG exam in Şanlıurfa province were used. The sample consisted of 64,852 students and 539 secondary schools in Şanlıurfa including all the secondary schools and their students participating in both TEOG exams in Şanlıurfa. The examination of all the schools' net average of the TEOG exams proved that they ranged from 25.24 to 69.52 for the former exam and from 27.70 to 70.21 for the latter one. As the two dataset of the study were not equated, the analyses were performed using Generalized Persistence [GP] model which is among the VAMs that do not assume equating. During the reporting procedure, numbers were used for schools instead of their names.

Generalized Persistence Model. As shown by Mariano et al. (2010), generalized persistence model is within the group of multivariable VAMs (McCaffrey et al., 2004). It is based on predicting school effect for more than one year simultaneously. Students' changing school or class during the year disrupts the clustered pattern desirable for the HLMs. The multivariate model that can be used in such more complex structures is described below:

Generalized persistence model assumes that the student's *t* year score depends on the overall *t* year mean for all the students, that teacher effect for current and previous years depends on the cumulative total, and that the test depends on the residual error of the students in the current year. If the term y_{it} is supposed to represent the student achievement of *i* in the year of *t*, the model for this score can be formulated as follows:

$$y_{it} = \mu_t + \left(\sum_{g=1}^{t} \sum_{j=1}^{J_g} \phi_{igj} \theta_{g[jt]} \right) + \varepsilon_{it}$$

In this equation, while the term μ_t represents the overall average for that year, the term ϕ_{igj} is entered to be 1 if the student is taught by the teacher (or school) *j* in the year of *g* or to be 0 otherwise. Therefore, the multiplication of $\phi_{igj}\theta_{g[jt]}$ demonstrates teacher or school effect for the current and previous years. Residual error is represented by ε_{it} in the equation. In this model, the term ε_{it} is assumed to have a normal distribution with zero average and unstructured covariance matrix ($\varepsilon_{it \sim MVN(0,\Sigma)}$). Mariano et al. (2010) indicate that the complete persistence and variable persistence models among the value-added assessment systems, which consider the persistence effect of the schools or teachers in previous years, can also be employed through this model. In this study, the model parameters and school effects of GP model were obtained by using GPVam package in R program (Karl, Yang, & Lohr, 2012). Maximum likelihood estimation method was used for the analyses. School effects (value-added scores) were estimated through the empirical best linear unbiased predictors (EBLUP).

Results

GPVam package has three options under general persistence models, GP (generalized persistence), VP (variable persistence) and CP (complete persistence) models. As the tests were not equated, the GP model was used in the study and the findings were reported. The parameter estimates obtained using the GP model were presented in Table 1. As a result of the analysed GP model, the AIC value, the relative fit measure, was estimated to be 500,137. In addition, the value-added scores of 539 schools were ascertained. They were calculated independently for the first and the second exams and shown in graphs in Figures 1 and 2.

Table 1. The Parameter Estimates of GP Model

	Coefficient	Standard Error
School effect for the first year	40.535	0.341
School effect for the second year	45.106	0.331

Table 2. The Covariance Coefficients of GP Model

	First year	Second year
First year	53.509	49.871
Second year	49.871	48.126

Note. The correlation coefficient of school effect between the first and second years was estimated to be .983.

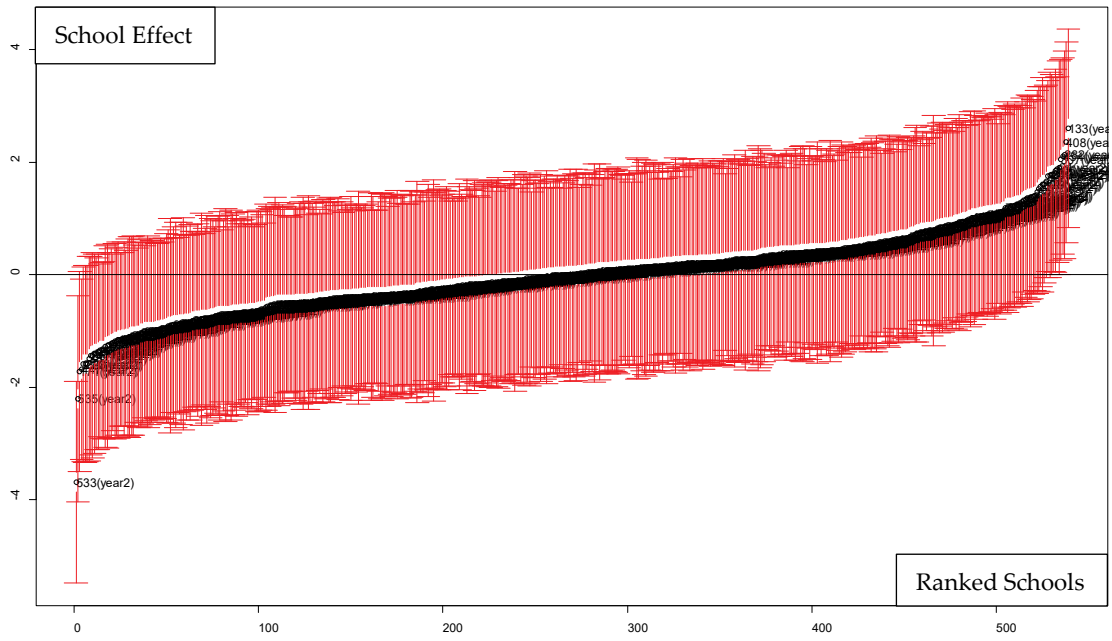

Figure 1. Value-Added Scores and Confidence Intervals for the First Year of Schools

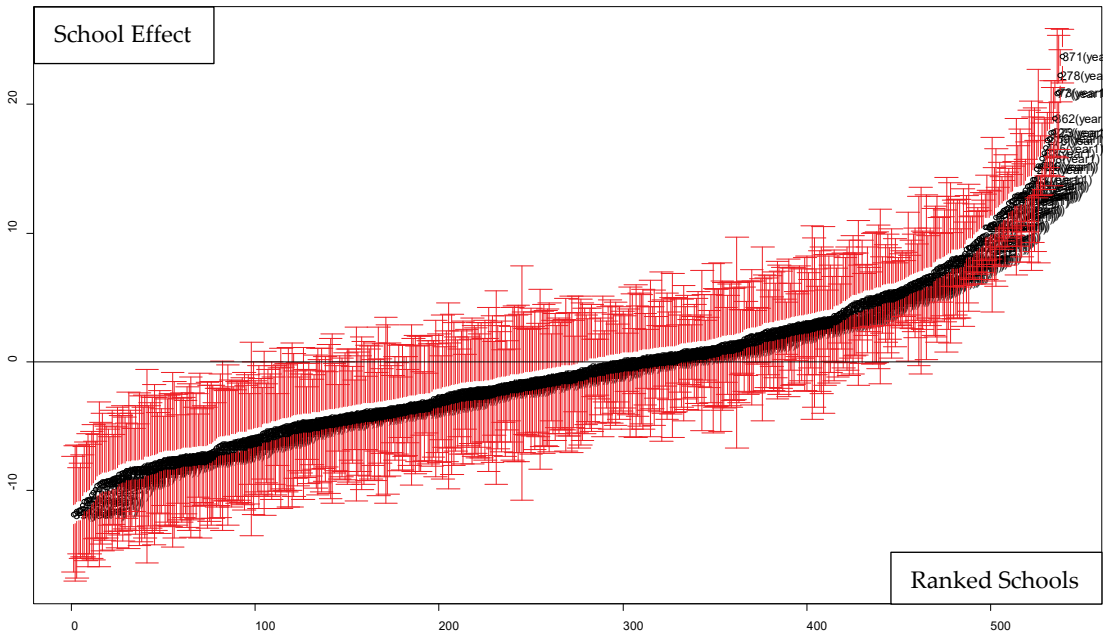

Figure 2. Value-Added Scores and Confidence Intervals for the Second Year of Schools

Value-added scores for the second test reflecting the change the students show from the former test to the latter one are what really need to be examined during the evaluation of school effectiveness. Table 3 and 4 presented the EBLUP values demonstrating the value-added scores obtained based on the increase and decrease of students' scores. They included the highest 20 and the lowest 20 schools with value-added scores. In addition, the net average of the TEOG exams ($\bar{X}_1$ = November, $\bar{X}_2$ = April) for these schools were also attached.

The examination of Table 3, presenting the scores of the highest 20 schools clarified that the three schools with the highest value-added scores were those with numbers 371, 278 and 73 respectively. It can also be observed that they were the ones with the high means among all schools in general. According to Table 3, it was clear that the schools with the highest means (the school with number 14 for the first exam and the school with number 77 for the second exam) did not have the highest value-added scores. It implied that the existing evaluation of school effectiveness only based on the exam averages produced divergent results from the value-added assessment method. While obtaining a value-added score for a specific school, the differentiation from the first exam to the second one for all students was taken into consideration. The ratings for the first and the second exams were also shown in Table 3 besides those according to the value-added scores. Although they were usually consistent with each other, schools with different ratings can also be identified. Correlation coefficients between the school ratings based on three distinct estimates were computed with Spearman Rho formula. While the correlation between the ratings depending on the value-added scores and the mean of the first exam was estimated to be .648, the correlation between the ratings according to the value-added scores and the mean of the second exam was .418.

Table 3. The Value-Added Scores and Means of the Highest 20 Schools

Rating	School Number	EBLUP*	Standard Error	$\bar{X}_1$	Rating ($\bar{X}_1$)	$\bar{X}_2$	Rating ($\bar{X}_2$)
1.	**371**	23.730	1.259	67.459	3.	68.071	4.
2.	**278**	22.210	1.239	65.465	5.	66.818	5.
3.	**73**	20.906	2.700	66.833	4.	68.646	3.
4.	77	20.847	3.041	67.708	2.	70.208	1.
5.	362	18.885	1.794	63.087	6.	63.556	14.
6.	273	17.825	2.108	61.679	7.	63.750	13.
7.	125	17.763	2.488	60.609	10.	66.707	6.
8.	359	17.319	1.709	60.941	9.	62.086	18.
9.	276	17.195	1.301	60.183	11.	61.687	19.
10.	405	16.617	1.331	59.121	14.	61.583	21.
11.	63	16.174	1.695	58.888	15.	61.673	20.
12.	258	15.759	1.681	57.501	20.	62.184	17.
13.	426	15.221	1.411	57.703	17.	60.198	24.
14.	484	15.218	4.567	69.524	1.	69.762	2.
15.	272	14.952	1.364	57.590	18.	59.658	28.
16.	237	14.143	3.376	59.305	13.	64.213	11.
17.	43	14.136	2.952	57.327	21.	63.794	12.
18.	199	13.650	3.991	61.666	8.	65.454	7.
19.	213	13.646	3.599	59.388	12.	64.477	9.
20.	351	13.605	2.120	57.561	19.	58.354	32.

*EBLUP = Empirical Best Linear Unbiased Predictors

The examination of Table 3, presenting the scores of the lowest 20 schools indicated that the three schools with the lowest value-added scores were those with numbers 26, 18 and 39 respectively. It can be observed that there were other schools with lower means. The difference between the value-added scores and the exam averages can be observed as in the highest schools. It implied that the available evaluation of school effectiveness only based on the exam averages created different results rather than the value-added assessment method. The ratings for the first and the second exams were also shown in Table 4 besides those according to the value-added scores. Although they were usually consistent with each other, there were also schools with different ratings. Correlation coefficients between the school ratings based on three distinct estimates were computed with Spearman Rho formula. While the correlation between the ratings depending on the value-added scores and the mean of the first exam was estimated to be .564, the correlation between the ratings according to the value-added scores and the mean of the second exam was .397.

Table 4. The Value-Added Scores and Means of the Lowest 20 Schools

Rating	School Number	EBLUP*	Standard Error	$\bar{X}_1$	Rating ($\bar{X}_1$)	$\bar{X}_2$	Rating ($\bar{X}_2$)
520.	190	-9.257	2.465	30.450	498.	33.452	515.
521.	275	-9.321	2.641	31.238	478.	31.690	528.
522.	300	-9.329	2.731	29.271	509.	33.880	507.
523.	137	-9.345	3.251	28.083	524.	33.241	521.
524.	25	-9.437	2.872	28.214	522.	34.464	500.
525.	311	-9.518	3.880	27.361	533.	29.375	536.
526.	185	-9.645	2.444	29.783	503.	33.333	517.
527.	437	-9.949	2.292	28.810	515.	34.166	502.
528.	70	-10.146	3.194	27.706	531.	31.547	530.

Table 4. Continued

Rating	School Number	EBLUP*	Standard Error	$\bar{X}_1$	Rating ($\bar{X}_1$)	$\bar{X}_2$	Rating ($\bar{X}_2$)
529.	138	-10.327	2.996	27.766	529.	31.933	526.
530.	4	-10.675	2.586	27.716	530.	32.725	524.
531.	10	-10.731	3.140	27.386	532.	30.477	533.
532.	224	-10.849	2.799	28.500	517.	30.361	534.
533.	395	-10.864	2.613	26.569	536.	33.578	513.
534.	495	-11.181	2.423	28.022	526.	31.590	529.
535.	314	-11.565	2.260	28.116	523.	31.053	532.
536.	389	-11.586	2.030	28.502	516.	31.248	531.
537.	**26**	-11.641	3.140	25.947	537.	29.696	535.
538.	**18**	-11.787	3.194	26.785	534.	27.698	239.
539.	**39**	-11.838	2.731	28.000	527.	28.567	538.

* EBLUP = Empirical Best Linear Unbiased Predictors

As a result, no problem has been experienced in the functioning of the system. In other words, it can be alleged that the value-added assessment system runs properly with TEOG data, which is a central government examination.

Conclusion and Discussion

The value-added assessment system among the most frequently used approaches to evaluate school effectiveness was introduced and implemented in this study through a sample data. The value-added assessment system estimates the value-added scores using the clustered data of teacher/student or school/student to attribute the improvement of students to the teachers or to the schools. The present study ranked the schools through the estimation of value-added scores based on the exam averages of the students who attended two distinct TEOG examinations in Şanlıurfa. Meanwhile, the students were also ranked merely based on the exam averages and it was concluded that this kind of rating greatly differed from the one based on value-added scores. The research findings pinpointed that the schools with high averages had superior value-added scores while those with low averages had inferior value-added scores to a great extent. Although the ratings seem consistent with each other, it should be noted that the value-added assessment approach produces more accurate results. Based on the findings and the studies in the relevant literature (Hanushek, 1972; McCaffrey et al., 2003; Sanders et al., 1997; Tekwe et al., 2004), it is conceived that the decisions on school performance according to the exam averages in a single attempt may lead to misjudgements. While the exclusion of school and student related variables is among the drawbacks of the traditional method on deciding whether the schools are successful or not on the basis of the exam averages, it seems possible to obtain more accurate results by controlling such kind of variables through the VAMs. As there were no additional variables for the students or schools in this study, model analyses were performed without predictive variables. It can be regarded as the limitation of the research.

The VAA system has the potential to be a strong alternative for the assessment of school effectiveness. Moreover, if teacher mobility is minimized, it can also play an important role in evaluating teacher effectiveness. It is hoped that the use of this assessment in a teacher evaluation system may eliminate the problem of "grading teachers" which has occupied the agenda of Turkey (Tedmem Report, 2018).

Given that the school administrators do not have concrete tools when evaluating teachers, the VAA system can provide administrators with the opportunity of objective evaluation. The drawbacks in practice can be eliminated through the VAA as it has been determined that some teachers do not find it appropriate to be observed and evaluated within the classroom by their superiors and the administrators have neglected teacher evaluation as they are not well-trained (Bozkuş, 2018).

According to the law "No Child Left Behind [NCLB]" enacted in the USA in 2001, the states are obliged to measure student achievement by applying standardized exams annually to the students studying at schools within their borders and improve those who are behind the anticipated level (McCaffrey et al., 2004). Since then, quite a lot of states have evaluated schools and teachers based on the longitudinal data of students through establishing their own value-added assessment systems. While the schools with satisfactory level of achievement are provided with incentives, those failed to progress as expected have been imposed on sanctions. The adoption of a similar practice in cities and schools in Turkey under the management of the Ministry of National Education may eliminate the difference between student achievement across the country. The present study sets a good example to test the feasibility of the VAA system in Turkish education context.

Many statistical models have been proposed in the literature to obtain the value-added scores to evaluate the performance of schools or teachers in the value-added assessment system. The American Statistical Association (2014) highlighted that the use of these models for the accountability of schools has been increasing in recent years, that the teachers and schools have been ranked on the basis of value-added assessments and that awards and sanctions have been placed in accordance with the ratings. The American Statistical Association (2014) and various researchers (Baker et al., 2010; Ballou & Springer, 2015; Beardsley, 2008; Goldhaber, 2015; Johnson, 2015) suggest that the results obtained from these models should not be used solely to make important decisions about schools and teachers. The researchers recommend that the value-added assessment approach should be used in school and teacher evaluation, not as the main decision-maker, but as a means of decision-making. In this regard, this study can be considered a proposal for the Ministry of National Education.

References

Adams, E. (1993). School's out!: New initiatives for British school grounds. *Children's Environments, 10*(2), 180-191.

American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Alexandria, VA.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., … & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper #278). Washington: Economic Policy Institute.

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher, 44*(2), 77-86.

Ballou, D., Sanders, W. L., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37-66.

Barnard, C. (1938). *The functions of the executive*. Cambridge, MA: Harward University Press.

Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher, 37*(2), 65-75.

Botha, R. (2010). School effectiveness: Conceptualising divergent assesment approaches. *South African Journal of Education, 30*(4), 605-620.

Bozkuş, K. (2018). *Öğretmenlerin mesleki gelişiminde dinamik yaklaşımın uygulanması: Bir eylem araştırması* (Unpublished doctoral dissertation). Anadolu University, Eskişehir.

Briggs, D. C., & Domingue, B. (2013). The gains from vertical scaling. *Journal of Educational and Behavioral Statistics, 38*(6), 551-576.

Briggs, D. C., & Weeks, J. P. (2009). The sensitivity of value-added modeling to the creation of a vertical score scale. *Education Finance and Policy, 4*(4), 384-414.

Creemers, B. P., & Kyriakides, L. (2008). *The dynammics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.

Doran, H., & Izumi, L. T. (2004). *Putting education to the test: A value-added model for California*. San Francisco, CA: Pacific Research Institute.

Doran, H. C., & Cohen, J. (2005). The confounding effect of linking bias on gains estimated from value-added models. In R. Lissitz (Ed.), *Value-added models in education: Theory and application* (pp. 80-110). Maple Grove, MN: JAM Press.

Drury, D., & Doran, H. C. (2003). The value of value-added analysis. *Policy Research Brief for National School Board Association, 3*(1), 1-4.

EARGED. (2010). *İlköğretim okulları başarı göstergeleri*. Ankara: MEB Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher, 44*(2), 87-95.

Hanushek, E. A. (1972). *Education and race.* Lexington, MA: Lexington Books.

Hesapçıoğlu, M. (2006). Eğitim kurumlarında kalite olgusu ve kalite güvence sistemleri. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi, 23*(1), 143-160.

Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school?. *Educational Researcher, 44*(2), 117-126.

Karip, E., & Köksal, K. (1996). Etkili eğitim sistemlerinin geliştirilmesi. *Eğitim Yönetimi, 2*(2), 245-257.

Karl, A. T., Yang, Y., & Lohr, S. (2012). GPvam: Maximum likelihood estimation of multiple membership mixed models used in value-added modeling. R Package Version 2.0-0 [Computer software]. Retrieved from https://cran.r-project.org/web/packages/GPvam/GPvam.pdf

Kelly, F. J. (1916). The Kansas Silent Reading Tests. *Journal of Educational Psychology, 7*(2), 63-80.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk, 1*(2), 147-169.

Mariano, L. T., McCaffrey, D. F., & Lockwood, J. R. (2010). A model for teacher effects from longitudinal data without assuming vertical scaling. *Journal of Educational and Behavioral Statistics, 35*(3), 253-279.

McCaffrey, D., Lockwood, J. R., Koretz, D., & Hamilton, L. (2003). *Evaluating value-added models for teacher accountability*. Washington, DC: RAND.

McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, *29*(1), 67-101.

Meyer, R. H., & Dokumaci, E. (2010). *Value-added models and the next generation of assessments*. Princeton, NJ: Center for K-12 Assessment & Performance Management. Retrieved from www.k12center.org/rsc/pdf/MeyerDokumaciPresenterSession4.pdf

Ministry of National Education. (2019). *Milli eğitim istatistikleri örgün eğitim*. Ankara: Milli Eğitim Bakanlığı. Retrieved from http://sgb.meb.gov.tr/meb_iys_dosyalar/2019_09/30102730_meb_istatistikleri_orgun_egitim_2018_2019.pdf

Murnane, R. J. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger Publishing Co.

Rowan, B. (1985). The assessment of school effectiveness. In R. Kyle (Ed.), *Reaching for excellence: An effective schools sourcebook.* Washington, DC: National Institute of Education.

Sanders, W. L. (2000). Value-added assessment from student achievement data: Opportunities and hurdles. *Journal of Personnel Evaluation in Education, 14*(4), 329-339.

Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Educational Assessment System (TVAAS): A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137-162), Thousand Oaks, CA: Corwin Press.

Tedmem Report. (2018). *Öğretmen performans değerlendirme ve aday öğretmenlik iş ve işlemleri yönetmelik taslağı üzerine değerlendirmeler.* Retrieved from https://tedmem.org/download/ogretmen-performans-degerlendirme-aday-ogretmenlik-is-islemleri-yonetmelik-taslagi-uzerine-degerlendirmeler?wpdmdl=2641

Tekwe, C. D., Carter, R. L., Ma, C., Algina, J., Lucas, M. E., Roth, J., & … Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11-36.

Temple, J. A., & Reynolds, A. J. (1999). School mobility and achievement: Longitudinal findings from an urban cohort. *Journal of School Psychology, 37*(4), 355-377.

Turhan, M., Şener, G., & Gündüzalp, S. (2017). Türkiye'de okul etkililiği araştırmalarına genel bir bakış. *Turkish Journal of Educational Studies, 4*(2), 103-151.

UNESCO. (2019). *Combining data on out-of-school children, completion and learning to offer a more comprehensive view on SDG 4* (Information Paper No. 61). Montreal, Quebec: United Nations Educational Scientific and Cultural ORganization. Retrieved from http://uis.unesco.org/sites/default/files/documents/ip61-combining-data-out-of-school-children-completion-learning-offer-more-comprehensive-view-sdg4.pdf

Yükçü, S., & Atağan, G. (2009). Etkinlik, etkililik ve verimlilik kavramlarının yarattığı karışıklık. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 23*(4), 1-13.