Students' Perspective on the Impact of English Teacher Development Programs on Teaching Quality in Indonesia

Azkiyah, S. N.[1]*, Doolaard, S.[2], Creemers, Bert P. M.[2] and Van der Werf, M. P. C. (Greetje)[2]

[1]*Faculty of Educational Sciences, Syarif Hidayatullah State Islamic University of Jakarta, Indonesia*
[2]*Groningen Institute for Educational Research, The University of Groningen, the Netherlands*

ABSTRACT

Finding effective methods for improving teaching quality is crucial because it has been found to be the most important factor in education. This paper compares two approaches of improving teaching quality through a one-year experimental study involving two experimental groups and one control group. The first is the use of education standards and the second is the use of education standards combined with a teacher development program. Teaching quality was measured in four variables: Building classroom as a learning environment (CLE), instruction, questioning, and orientation. In this study, 1255 students and 45 teachers from 43 junior secondary schools in two provinces of Indonesia voluntarily participated. Multilevel modelling was employed and the results indicate that both interventions have significant effects on the outcomes. However, as expected, the second intervention is significant in all variables and has larger effects, whereas the first intervention is significant only in two variables: CLE and questioning. The findings of this research imply the need to have clear and concrete education standards and to enhance these education standards with training sessions to facilitate better teaching quality.

Keywords: Education standards, teaching quality, teacher professional development program, dynamic model of educational effectiveness research

INTRODUCTION

It is widely accepted that teachers play important roles in improving learning outcomes. Therefore, various approaches have been introduced and implemented to improve teacher quality. In this paper, two

approaches, namely education standards and educational effectiveness research (EER) are compared through a one-year experimental study. Education standards have been argued to serve as guidance for improvement purposes, which is the focus of this paper. The argument is that standards define goals and provide concrete direction to different stakeholders, such as principals, teachers and administrative assistants to provide learning opportunities and improve outcomes for all students regardless of their background (Dowson, McInerney, & Van Etten, 2007; Neumann, Fischer, & Kauertz, 2010; Schmidt, Houang, & Shakrani, 2009, Stosich, 2016, Volante, 2012). Agreement on a shared set of specific, clearer and higher goals and ways to accomplish them is found to improve organisational capacity and to have better planning and actions (Kawamato & Greenes, 2014). In this way, both excellence and equity are argued to be simultaneously addressed. .

For teachers, education standards provide direction by setting clear indication on what students should learn and be able to do (Choi, de Vries, & Kim, 2009; Dowson, et al., 2007; Marzano & Kendall, 1996; National Research Council [NRC], 2001; Neumann et al. 2010; Ravitch, 1995; Schmoker & Marzano, 1999). In this case, teachers are required to deepen their knowledge and skills related to their subject matter and incorporate various teaching strategies to meet different needs of different students, which in turn are expected to improve learning outcomes for all students

(Chambers & Dean, 2000; McClure, 2005; Stosich, 2016).

The standards are also argued to promote school accountability whereby schools have to report their results to the public and receive the consequences accordingly (Raizen, 1998; Ravitch, 1995). In this respect, high-stakes testing has been widely developed to monitor performance (Hamilton, Stecher, & Yuan, 2008). However, the assessment is more summative, where results are used mainly for accountability purposes and not to provide feedback for teachers to improve their instruction (Snow-Renner, 2001). In addition, little is known about the results. Several existing studies, which are dominated by those in the United States, have identified a large variation among different states despite some progress (Hanushek, Peterson, & Woessmann, 2012) and performance gaps among different states and between the white and black students (e.g. Dowson et al., 2007; Hanushek et al., 2012; US Department of Education, 2008, 2015).

Indonesia, as the focus of the study, has a history of low student achievement in both national and international measurements (Jazadi, 2003; Kompas, 2010; Mohandas, 2004; Nurweni & Read, 1999; Setiogi, 2003; Tilaar, 1992). Following the movement of education standards in other countries, the Indonesian government established formal education standards to improve education in 2005. Schools are expected to start implementing them seven years afterwards at the latest (Pemerintah Indonesia, 2005). To

date, as in other countries, few experimental studies have been conducted to examine the effectiveness of the standards. Referring also the findings in the US, thus, not much could be learnt from the standard movement; there are, indeed, missing gaps that need to be fulfilled.

Two problems may contribute to the ineffectiveness of the standards. Firstly, research has found that standard documents have been largely written in a broad and global language (e.g. Choi, de Vries, & Kim, 2009; Dowson et al. 2007; Hammer, 1998) and have been criticised for being vague and nebulous and having insufficient grounding in their content (Gandal, 1996; Finn, Petrili, & Vanourek, 1998). The standards usually consist of lists of topics and there is lack of explicit guidance for teachers about instructional strategies (Hill, 2001). Secondly, teachers face problems dealing with the standards (Gandal, 1996; Hammer, 1998; NRC, 2001, Stosich, 2016), which clearly implies that teachers are not provided with appropriate assistance.

On the other hand, educational effectiveness research (EER) has provided empirical evidence on factors related to student performance, which are useful to guide educational improvement and to prioritise actions. There have been different lists of factors situated at different levels, namely context/system, school, classroom/teacher and student levels. However, the findings suggest a consensus that high expectations, curriculum quality/opportunity to learn, school climate, classroom climate, effective learning time/classroom management, structured instruction, and frequent evaluation are found to be strongly related to student outcome (Azigwe, 2016; Creemers & Kyriakides, 2008; Creemers & Reezigt, 1996; Ehren, 2015; Muijs & Reynolds, 2000, 2011; Reynolds et al., 2014).

Another important finding of EER is the fact that teacher or classroom factors, which in some studies are referred to as teacher instructional roles, have been found to be superior compared to factors at other levels (e.g. Creemers, 1994, Darling-Hammond, 1997; Doolaard, 1999; Goldhaber, 2015; Harris & Muijs, 2005; Luyten & Snijders, 1996; Marzano, 2007; Scheerens, 2013; Van Der Werf, Creemers, De Jong, & Klaver, 2000).

There have been models of EER developed by researchers in the field, one of which is the dynamic model of educational effectiveness research (Creemers & Kyriakides, 2008).

As described in Figure 1, it has four levels, that are, context/national policy level, school level, teacher/classroom level, and student level. The model emphasises the teacher/classroom level and expects the above levels, namely context and school levels to provide necessary conditions for effectiveness of the classroom level.
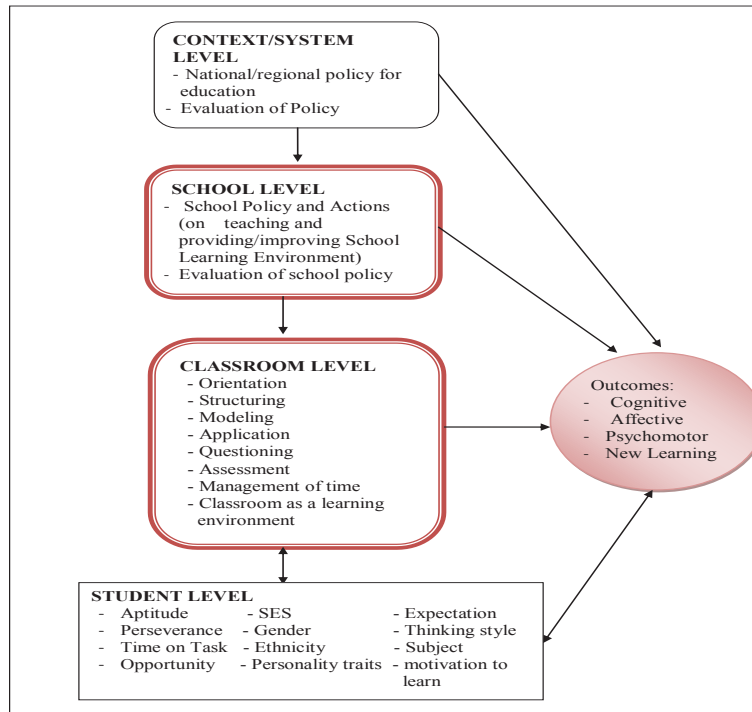
Figure 1. The dynamic model of Educational Effectiveness Research (Creemers & Kyriakides, 2008)

Responding to the criticism that existing effectiveness models do not explicitly refer to the measurement of each effectiveness factor, the model proposes five dimensions to measure the effectiveness of factors in each level. Creemers and Kyriakides (2008) argue that the dimensions could provide a better picture of effectiveness enhancing factors and hence, more specific strategies for improving educational practice could be established. These dimensions are frequency, focus, stage, quality and differentiation. Frequency refers to the quantity of activities associated with effectiveness factors, whereas focus is dealing with the specificity of the activity in relation with the goals of the activity. Effectiveness factors can

happen in different periods and therefore the dimension of stage looks at the period at which the activities take place. Quality looks at the properties of the activities or whether they are supported by literature or whether students understand and can follow the activities. Finally, differentiation is concerned with the diversity of students and teacher actions to help all students have the opportunity to learn.

With respect to the classroom level of the model, which is the main concern of the model and also this paper, Creemers and Kyriakides (2008) refer to observable teacher instructional roles and relate them to student outcomes. These factors include: (1) orientation; (2) structuring; (3) questioning;

(4) teaching modelling; (5) application; (6) management of time; (7) teacher role in making classroom a learning environment; and (8) classroom assessment. Table 1 describes the main elements of each factor. An experimental study by Antoniou (2009) using classroom factors of the dynamic model found an increase of teaching quality and student performance and therefore, the model is adopted in this paper.

In summary, the standards serve as guidance in setting educational goals and expectations, whereas the results of EER provide empirically validated principles on what to prioritise in the teacher development program in order to achieve the goals.

Taking both the arguments of education standards and the empirical findings of EER in serving frameworks to improve teacher quality, this paper attempts to compare two approaches: the use of standards and the use of standards combined with a teacher development program. However, as the case in other countries, the Indonesian education standards have been written in broad and general language. Therefore, this paper proposes an elaborated standards document for English in junior high schools as the context of the study, which is intended to further explain the competencies mentioned in the standards.

Table 1
The main elements of each teacher factor included in the dynamic model (Creemers & Kyriakides, 2008)

1)	Orientation	a)	Providing the objectives/goals for which a specific task/lesson/series of lessons take(s) place
		b)	Challenging students to identify the reason for which an activity takes place in the lesson
2)	Structuring	a)	Beginning with overviews and/or review of objectives
		b)	Outlining the content to be covered and signalling transitions between lesson parts
		c)	Calling attention and reviewing main ideas
3)	Questioning	a)	Raising different types of questions (i.e. process and product) at appropriate difficulty level
		b)	Giving time to students to respond
		c)	Dealing with student responses
4)	Teaching modelling	a)	Encouraging students to use problem solving strategies presented by the teacher or other classmates
		b)	Inviting students develop strategies
		c)	Promoting the idea of modelling
5)	Application	a)	Using seatwork or small group tasks in order to provide the necessary practice and application opportunities
		b)	Using application tasks as starting point for the next step of teaching and learning

Table 1 *(continue)*

6)	The classroom as a learning environment	a)	Establishing on task behaviour through the interactions they promote (i.e. teacher-student and student-student interactions)
		b)	Dealing with classroom disorder and student competition by establishing rules, persuading students to respect them and using the rules
7)	Management of time	a)	Organising the classroom environment and maximising engagement rates
8)	Assessment	a)	Using appropriate techniques to collect data on student knowledge and skills
		b)	Analysing data in order to identify student needs and report the results to students and parents
		c)	Evaluating teachers' own practices.

RESEARCH QUESTIONS AND HYPOTHESES

This paper raises the following two research questions:

1. Do the (elaborated) standards only improve teachers' teaching quality?

2. Does the combination of the (elaborated) standards and a teacher professional development program (TPD) improve teachers' teaching quality more?

In line with these questions, two hypotheses are examined:

H1: teaching quality will be better when the education standards are further elaborated.

H2: teaching quality will be much better when the elaborated standards document is combined with a TPD.

EER, especially the classroom factors of the dynamic model of educational effectiveness research, was used to design the TPD. Two intervention groups were set up. The first group worked with the elaborated standards while the second group combined the elaborated standards document with a teacher improvement program.

METHODS

Research Design

The study was an experimental study in which two interventions were deliberately introduced to observe their effects (Field, 2009). The participants voluntarily participated, and as described in Table 2, were randomly assigned into three groups.

Table 2
The groups and the intervention

Group		Intervention
1.	Experimental 1	The elaborated standards document; the teachers were free to develop their own strategies in implementing and achieving the standards
2.	Experimental 2	The elaborated standards document and a teacher development program (teacher training)
3.	Control group	No intervention, teachers used the standard document available from the government

Participants

The population of the study was a private Madrasah Tsanawiyah in the provinces of DKI and Banten. Indonesia has a dual schooling system: general school, which is managed and supervised by the Ministry of National Education (MONE) and *madrasah*, which is under the authority of the Ministry of Religious Affairs (MORA). Students in *madrasah* follow the same national curriculum complied with by those in general schools but they have additional subjects for religions and therefore, have longer hours of schooling (Departemen Pendidikan Nasional, 2003). This study was specifically conducted for Madrasah Tsanawiyah, a junior secondary school level (12 to 13, 15 to 16 years old) due to several considerations. Most madrasah schools are private, small, and attended by students from low income families and generally are of lower quality compared to general schools (Asian Development Bank, 2006, Centre for Excellency and Quality Development Assurance [CEQDA], 2007). Therefore, improvement in this type of school is urgently required. English, especially reading comprehension, was selected since English is one of the subjects tested nationally through the national exam and students' attainment on this subject has been the lowest throughout the years (Badan Standar Nasional Pendidikan, 2010).

Sampling was carried out at the school level, yet the focus of this study was on the teacher and student levels. In each of the two provinces three municipalities or districts with the highest number of madrasahs were selected. Another criterion concerned school size; schools with enrolment of above 100 students were selected. Information on the research project and invitation to participate was sent to more than 200 schools in the selected areas. Until the deadline, 57 schools (32% accredited A, 68% accredited B), with a total of 59 teachers (M = 44%, F = 56%) and 2,431 students (M = 48, 5%, F = 51, 5%) volunteered to participate. In this paper, referring to the analysis, which requires the participants to participate in both measurements, only 45 teachers and 1255 students are retained.

Some background characteristics at both teacher and student levels were collected in this study. At the teacher level, the characteristics included gender, teachers' qualification, , years of teaching experience, and age while those at the student level consisted of gender, fathers' and mothers' education and fathers' and mothers' jobs. These characteristics are displayed in Table 3 and 4 respectively, which generally show no significant differences among the three groups.

Table 3
The background characteristics of teachers according to group

Characteristic	Distribution (%)		
	Exp 1	Exp 2	Cont
Teacher gender			
1. Female	50	74	47
2. Male	50	26	53
Teacher degree			
1. Diploma	13	21	18
2. Bachelor	81	74	77
3. Master	6	5	6
Teacher major			
1. English	75	84	82
2. Non-English	19	11	18
3. Missing/unknown	6	5	0
Teacher age			
1. <= 30 years	50	47	47
2. 31 - 40 years	25	42	18
3. 41 -5 0 years	25	11	29
4. Missing/unknown			6
Teaching experience			
1. <= 5 years	44	42	31
2. 6 - 10 years	31	32	31
3. > 10 years	25	26	38

Note: Exp 1 refers to experimental group 1, Exp 2 to experimental group 2, and Cont to the control group

Table 4
The background characteristics of students according to group

Characteristic	Distribution (%)		
	Exp 1	Exp 2	Cont
Student gender			
1. Female	50	55	52
2. Male	50	45	48
Father's education			
1. Primary	22	13	24
2. JSS	24	23	25
3. SSS	33	39	26
4. University	6	10	4
Missing	15	16	21
Mother's education			
1. Primary	32	19	35
2. JSS	23	27	19
3. SSS	26	28	20
4. University	4	7	4
Missing	16	18	22
Father's Job			
1. Labour and Farmer	32	26	38
2. Small business	40	34	38
3. Professional	17	25	10
Missing	12	15	14
Mother's Job			
1. Housewife	82	72	75
2. Labour and Farmer	2	3	4
3. Small business	6	6	6
4. Professional	5	11	4
Missing	5	9	11

Note: Exp 1 refers to the experimental group one, Exp 2 to experimental group 2, and Cont to the control group

Research Instrument

The main variable in this study is teaching quality, which was measured through classroom observation by independent observers and a student questionnaire. This paper specifically reports the results

of the student questionnaire, which was distributed two times (mid and end points) and constructed based on the classroom factors of the dynamic model (Creemers & Kyriakides, 2008). The students filled in the questionnaire merely based on their perception since there was no training for them on how to fill in the questionnaire nor any interview to confirm their perception. They rated their teachers twice, in the middle and end of intervention Due to technical problems, students were unable to rate their teachers prior to the intervention. However, based on the data of the observer, no difference in teaching quality (F {2, 49} = .554, p = .578) was observed among the three groups.

During the pilot study (> 300 students), 40 items representing the classroom factors of the dynamic model were included and questions were based on a five-point unidirectional scale ranging from "never" to "a great deal". In the exploratory factor analysis, four scales (32 items, α ranged from .71 to .85) were resulted and retained in the present study. These include orientation, instruction (the items refer to structuring, modelling, and application), questioning, and creating classroom as a learning environment (CLE). In both measurements in the present study, the (α) reliability was equal or above .80 except CLE in the second measurement (.71). In all measures, several items were reversed to assure that all items go in the same direction.

Interventions in the Study

Prior to the study, an analysis of the Indonesian government standards (of content) of English (reading comprehension) was carried out to analyse the clarity of the terms. It revealed that they were broad and general, and therefore, elaborated standards were developed to make them more concrete and specific. Both experimental groups attended a workshop to discuss the document. The activities included discussion on specific reading skills, analysis of the government standards to get an agreement on the proposed elaborated standards, and analysis and development of questions/tasks in accordance with the elaborated standards. Teachers in experimental group 1 were free to choose strategies to implement the elaborated standards.

Teachers in experimental group 2 were further assisted in a TPD program. In addition to the elaborated standards document, another document on effective teaching based on the classroom factors of the dynamic model was developed for experimental group 2. The TPD consisted of six meetings and the sessions were a combination of brainstorming, lecture, discussion, and group work in which they presented the result. The topics were: (1) building student interest and motivation through orientation and structuring; (2) teaching strategies [modelling] and the development of tasks [application] and questioning (for skills/competencies under

"understanding meaning"); (3) teaching strategies [modelling] and the development of tasks [application] and questioning (for skills/competencies under "responding meaning"); (4) developing lesson plan 1 (theoretical); (5) developing lesson plan 2 (practical); and (6) peer teaching. In the visit during the observation, feedback was provided by the observer to teachers in experimental group 2.

Data Analysis

The data were analysed both descriptively and statistically. The descriptive analysis refers to the mean score of each group in the four outcome variables (orientation, instruction, questioning, and CLE). These mean scores of each scale were computed and analysed using multilevel modelling (using MLwiN; Rasbash, Charlton, Browne, Healy, & Cameron, 2005) to investigate the effects of the interventions. The data were nested in three levels: student (level 1), school (level 2), and group (level 3) in which two levels were included and group (indicating intervention) was used as predictors. The first step was an empty

model to understand the variance at the student and school levels. The next step added the result of measurement one to see its effect. The third step included not only the result of measurement one but also the second intervention (experimental group two) as the predictors and regarded the first intervention (experimental group one) and the control group as the control group to examine the effect of the second intervention on the outcome variables. However, in this model it is not known whether the first intervention contributed to the outcome variables. Finally, the last model compared the effects of both interventions.

RESULTS

Descriptive: General Picture of Teaching Quality

The descriptive finding presents the mean score of each group in all four outcome variables in both measurement one and two in order to provide general picture of teaching quality. In general, students perceived good teaching quality of their teachers in both measurement 1 and 2.

Table 5
Descriptive statistics

Outcome Variable/Group	Measurement 1			Measurement 2		
	N	Mean	SD	N	Mean	SD
CLE						
Experimental Group 1	422	3.22	.56	426	3.31	.57
Experimental Group 2	491	3.29	.60	494	3.41	.58
Control Group	333	3.16	.50	334	3.18	.64
Excluded	9			1		
Total	1255	3.24	.55	1254	3.32	.60

Table 5 *(continue)*

Outcome Variable/Group	Measurement 1			Measurement 2		
	N	Mean	SD	N	Mean	SD
Instruction						
Experimental Group 1	421	3.22	.61	425	3.36	.66
Experimental Group 2	491	3.23	.65	493	3.40	.72
Control Group	328	3.17	.62	335	3.26	.81
Excluded	15			2		
Total	1240	3.21	.63	1253	3.35	.72
Questioning						
Experimental Group 1	423	3.38	.59	426	3.41	.69
Experimental Group 2	492	3.42	.65	494	3.45	.69
Control Group	335	3.96	.54	335	3.22	.78
Excluded	5			0		
Total	1250	3.40	.60	1255	3.78	.726
Orientation						
Experimental Group 1	425	3.44	.72	426	3.43	1.03
Experimental Group 2	494	3.49	.78	294	3.54	.86
Control Group	334	3.47	.70	335	3.33	.88
Excluded	2			0		
Total	1253	3.47	.74	1255	3.40	5.93

Table 5 indicates that students in all groups started to rate their teacher high (above 3 from 1-5 scales) in measurement one. In all the four outcome variables, teachers in experimental group 2 were rated higher than those in the other two groups except questioning, which was rated higher by the control group. Similarly, the standard deviation is also highest in experimental group 2, indicating more variance of teaching quality in the group. It is important to note that the data in this measurement were collected in the middle of the intervention. Due to some problems, the data at the beginning of the intervention could not be collected. However, as previously noted, no significant difference, $F (2, 49) = .554$, $p =$ .578, was found in the data from independent observers collected at the beginning of the intervention.

Compared to the first measurement, students in experimental group 1 and 2 rated their teachers higher in all four variables in the second measurement, that is, there was a gain from the first to the second measurement. On the other hand, the control group had a decrease in questioning and orientation. Teachers in experimental group 2 were again rated higher than the other two groups, but this time in all variables. Overall, there was an increase in all four variables from measurement one to measurement two, except orientation, from 3.47 in the first measurement to 3.40 in the

second one. In short, the above descriptive statistics showed changes and variation of teaching quality as perceived by students, whose effects were further investigated through multilevel modelling analysis.

The Result of Multilevel Modelling Analysis: The Effect of Interventions

As depicted in Table 6, four models were tested, whose results also showed variance at both student and school levels in all variables. The empty model shows that the differences in all variables are bigger at the student than at school level, as is common in this type of study. In order to have a better estimate of the effect of the interventions, the subsequent analysis determined the results of the first measurement. Model 1 indicates that the first measurement in all variables has significant effect ($p < .01$) and is positively related to the results of the second measurement. Thus, students who rated high in the first measurement also rated their teachers high in the second one. The last model clearly shows that the first intervention has significant effects on two variables: CLE and questioning.

Table 6

The results of multilevel analysis explaining variation of teaching quality and the effects of interventions

	Empty Model	Model 1 (+ measurement 1)	Model 2 (+ intervention 2)	Model 3 (+ both interventions)
	Creating Classroom as a Learning Environment			
Fixed Part				
Constant	3.31 (.03)	2.54 (.10)	2.50 (.10)	2.43 (.10)
CLE (measure 1)		.24 (.03) ***	.23 (.03) ***	.23 (.03)
Intervention 1 (group 1)				.13 (.06)**
Intervention 2 (group 2)			.14 (.05) **	.21 (.05)***
Random Part				
School level	.03 (.01)	.02 (.01)	.01 (.00)	.01 (.00)
Student level	.34 (.01)	.33 (.01)	.33 (.01)	.33 (.01)
Deviance	2.244.363	2.175.734	2.168.675	2.163.489
Decrease in deviance		68.629	7.059	5.186
Variance explained		.052	.017	.009
Total variance explained			.068	
	Instruction			
Fixed Part				
Constant	3.35(.03)	2.60 (.11)	2.66 (.11)	2.61 (.12)
Instruction (measure 1)		.20 (.03) ***	.22 (.03)***	.20 (.03)***
Intervention 1 (group 1)				.10 (.07)
Intervention 2 (group 2)			.08 (.06)	.13 (.07)*

Table 6 *(continue)*

	Empty Model	Model 1 (+ measurement 1)	Model 2 (+ intervention 2)	Model 3 (+ both interventions)
Random Part				
School level	.02 (.01)	.02 (.01)	.02 (.01)	.02 (.02)
Student level	.50 (.02)	.49 (.02)	.49 (.02)	.49 (.02)
Deviance	2.730.265	2.666.475	2.664.750	2.662.902
Decrease in deviance		63.790	1.725	1.848
Variance explained		.028	.004	.004
Total variance explained			.032	
Questioning				
Fixed Part				
Constant	3.36 (.03)	2.60 (.01)	2.55 (.12)	2.44 (.12)
Questioning (measure 1)		.22 (.03) ***	.22 (.03)***	.22 (.03)***
Intervention 1 (group 1)				.19 (.07)**
Intervention 2 (group 2)			.12 (.06) **	.22 (.07)***
Random Part				
School level	.03 (.01)	.03 (.01)	.02 (.01)	.02 (.01)
Student level	.49 (.02)	.48 (.02)	.48 (.02)	.48 (.02)
Deviance	2.723.683	2.670.920	2.667.196	2.660.980
Decrease in deviance		52.763	3. 724	6.216
Variance explained		.038	0.008	0.012
Total variance explained			0.045	
Orientation				
Fixed Part				
Constant	3.44 (.04)	2.92 (.13)	2.863 (.13)	2.82 (.14)
Orientation (measure 1)		.15 (.04) ***	.15 (.04)***	.15 (.04)***
Intervention 1 (group 1)				.08 (.10)
Intervention 2 (group 2)			.15 (.08) **	.20 (.09)***
Random Part				
School level	.04 (.02)	.04 (.01)	.03 (.01)	.03 (.01)
Student level	.82 (.03)	.81 (.03)	.81 (.03)	.81 (.03)
Deviance	3.358.162	3.322.524	3.318.744	3.318.016
Decrease in deviance		35.638	3.780	.728
Variance explained		0.027	0.007	0.001
Total variance explained			0.034	

*** p < .01; ** p < .05; * p < .10

Model 2 demonstrates that experimental group 2 has significant effect in all variables at $p < .05$, compared to other groups, except instruction. On this variable, the effect is significant only at $p < .10$ (one tailed). However, Model 3 gives a better estimate: it is significant at $p = .058$. In addition, it is interesting to highlight that the second intervention explains nearly 2% of the variance in building classroom as a learning environment as indicated in Model 2. However, with respect to instruction, Model 2 shows that it explains almost nothing (.004) and the effect is less significant ($p < .10$, 1 tailed). Nevertheless, the variance at the student level in this model remains the same as that in Model 1, indicating that the variance is due to the differences at the school level, and thus the intervention plays a role. Moreover, Model 3 provides supporting information that the second intervention has significant effect at $p = .058$.

Overall, the findings, especially as described by the estimates in Model 3, clearly suggest that both interventions have effect but the second one has larger effects, not only in terms of the coefficient but also in the aspects. Concerning the aspects, experimental group 2 is proven to have significant effects in all four outcome variables, whereas in experimental group 1, there were significant effects in only two variables, namely, CLE and questioning.

DISCUSSION

The study examined the impact of two approaches of improving teaching quality as perceived by students. Generally, students reported high ratings in all four variables, which suggest that Indonesian teachers are considered to have relatively good teaching quality, as perceived by their students. However, this finding is contradictory to previous studies. Using classroom observation, Utomo (2005) described that in general, classrooms were characterised by a didactic, whole-class style of teaching. He, further explained that teachers paid little attention to children's needs as individual learners and little recognition that children enter school with ideas, opinions and conceptions about their world. Similarly, Kaluge, Setiasih, and Tjahjono (2004) indicated that teachers in general are unable to create active, joyful and effective learning environment.

This finding could be influenced by the Indonesian culture in which the teaching profession is normally regarded as a high and respected occupation and therefore, teachers benefit from being respected by students and society in general (Maulana, Opdenakker, Den Brok, & Bosker, 2011). Nevertheless, it is important to note that this study found significant effects of the second intervention, which is the main concern in the study, in all variables. Thus, teachers in experimental group 2 were perceived to have better performance by their students after one school year of intervention compared to teachers in the other two groups.

Furthermore, it is also crucial to show that the study found significant effects ($p < .05$) of the first intervention on two variables: building classroom as a learning

environment and questioning. However, a bigger effect size of the second approach could be inferred from its larger estimates compared to the first one. Therefore, the hypothesis built in this paper is satisfied. The use of clear and concrete education standards lead to better teaching quality compared to the general and broad education standards and the use of standards combined with teacher development program prove to outperform the other approaches in improving teaching quality.

It is also interesting to explore further the effects of both interventions on different variables. Concerning building classroom as a learning environment, the second intervention has significant effect at $p = .000$ and explains 2% of the existing variance, which is relatively high considering the fact that school level normally explains around 10% with many variables. In this variable, the first intervention has a significant effect at $p = .002$. Nevertheless, this result shows that teachers in both interventions have improved their skills, for instance, in building interactions among the members of the class and managing classroom disorder.

On the other hand, the results described in Table 6 indicate that even teachers in experimental group 2, to some extent, face difficulties to provide proper instruction to students. Model 2 explains that the effect of the second intervention is less significant ($p = .10$, 1 tailed and .058 in Model 2 and 3, respectively) compared to its effect in the other variables. Likewise, Model 3, which is considered as the best model, finds no significant effect of the first intervention

in instruction. This finding suggests that teachers may even forget to briefly explain the series of activities students will carry out and have more difficulties in presenting strategies or models of learning, and providing appropriate and enough application activities for their students. This result is not surprising as previous studies indicated similar findings, for instance, teachers are text-book oriented (Utomo, 2005), whereas some (English) text-books have been criticised for not enabling learners to achieve the goals of using the language and not providing further explanation on why and for what purposes students should do the exercises (Priyanto, 2009). It is also possible that instruction contains more difficult teaching skills.

With respect to questioning, the results indicate that teachers in both experimental groups have managed to raise various questions with more or even better hints and feedback to their students as both interventions have significant effect ($p < .01$ for the second intervention and $p < .05$ for the first one). Finally, the second intervention has significant effect on orientation ($p < .01$, model 2 and 3) but the first intervention does not. As previously mentioned, orientation is a very important aspect because effective teachers have been found to provide orientation activities to encourage students' motivation to learn, and this study has shown that the second intervention has succeeded in improving teaching skills in this aspect.

Finally, the teaching quality improvement shown in this study supports

previous studies such as by Avalos (2011) and Antoniou and Kyriakides (2013). Reviewing publications in the journal, *Teaching and Teacher Education* over ten years (2000 – 2011), Avalos concludes that when teachers participate in teacher development programs, improvement in teachers' knowledge and practice is observed. Similarly, the study of Antoniou and Kyriakides show links between teacher professional development and improvement in teaching skill.

CONCLUSION

This study compared two approaches of improving English teachers' teaching quality. The first was the use of elaborated education standards and the second was the use of elaborated education standards combined with a teacher development program, where the classroom factors of the dynamic model was used as the framework in designing the program. The findings of this study suggest two important conclusions. Firstly, education standards produce better teaching quality when the standards are written in concrete and clear language. Secondly, teachers' teaching quality is better improved when elaborated education standards are enhanced with a teacher development program or training sessions.

Thus, the findings of this study imply the need to provide teachers with a teacher development program. Furthermore, practically this study offers both theory-driven and evidence-based principles on the aspects and strategies of improving teaching quality. Theoretically, the study provides empirical evidence on the necessity of teacher development program next to the education standards and provides ground for further development of the dynamic model, especially concerning the use of standards.

Nevertheless, this study should also be considered from the perspective of some limitations. Firstly, this study was limited only to one type of schooling in Indonesia and one subject, of which results might not be generalisable to other contexts. Secondly, a previous study (Antoniou, 2009) revealed that teaching skills in the dynamic model could be classified into five stages, progressing from the easier to more difficult. When teachers are found in stage one, they are able to concentrate their improvement in skills in the second stage. Due to inevitable reasons, this study did not incorporate these stages. When these stages are included in a follow-up study, it is possible to see more comprehensive results.

REFERENCES

ADB. (2006). *Indonesia: Madrasah education development project.* Asian Development Bank. Project Number: 37475-01. Jakarta: SMEC International Pty Ltd.

Antoniou, P. (2009). *Using the dynamic model of educational effectiveness to improve teaching practice: Building an evaluation model to test the impact of teacher professional development programs.* (Unpublished doctoral thesis). University of Cyprus, Cyprus.

Antoniou, P., & Kyriakides, L. (2013). A dynamic integrated approach to teacher professional development: Impact and sustainability of the effects on improving teacher behaviour and student outcomes. *Teaching and Teacher Education, 29*, 1-12.

Antoniou, P., Demetriou, D., & Kyriakides, L. (2006). Towards a dynamic model of educational effectiveness. A meta-analysis of studies investigating the impact of school factors on student achievement gains. In *Proceedings of the IX Conference of the Cyprus Pedagogical Association*. Nicosia.

Avalos, B. (2011). Teacher professional development in teaching education over ten years. *Teaching and Teacher Education, 27*(1), 10-20.

Azigwe, J. B. (2016). Using comparative international studies for modeling educational effectiveness: A secondary analysis of PISA-2009 study. *Journal of Education and Practice, 7*(18), 199-209.

Brophy, J., & Good, T. L. (1986). Teacher behaviour and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd Ed., pp. 328-375). New York: MacMillan.

BSNP. (2010). *Statistik.* Badan Standar Nasional Pendidikan. Retrieved from http://bsnp-indonesia.org/id/?page_id=39/

CEQDA. (2007). *Laporan dampak kebijakan akreditasi, BOS, dan sertifikasi terhadap masa depan madrasah di DKI Jakarta.* Kerjasama LAPIS dan CEQDA UIN Jakarta.

Chambers, L., & Dean, C. (2000). *Noteworthy perspectives on implementing standards-based education.* Washington D.C.: Mid-Continent Research for Education, and Learning.

Choi, D., de Vries, H. J., & Kim, D. (2009). *Standards education policy development: Observations based on APEC research.* Rotterdam: Erasmus University Rotterdam Host.

Creemers, B. P. M. (1994). *The effective classroom.* London: Cassel.

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

Creemers, B. P. M., & Reezigt, G. J. (1996). School level conditions affecting the effectiveness of instruction. *School Effectiveness and School Improvement, 7*(3), 197-228.

Darling-Hammond, L. (1997). *Doing what matters most: Investing in quality teaching.* New York: National Commission on Teaching and America's Future.

DE. (2008). *A nation accountable: Twenty-five years after "A nation at risk".* Washington D.C.: US Department of Education.

DE. (2015). *School composition and black-white achievement gap.* The US Department of Education. Retrieved from https://nces.ed.gov/nationsreportcard/subject/studies/pdf/school_composition_and_the_bw_achievement_gap_2015.pdf

Departemen Pendidikan Nasional. (2003). *Undang-Undang Nomor 20 Tahun 2003, Tentang Sistem Pendidikan Nasional.* Jakarta: Depdiknas.

De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the 1st year of junior general education in the Netherlands. *School Effectiveness and School Improvement, 15*(1), 3-31.

Dowson, M., McInerney, D. M., & Van Etten, S. (2007). *The state of play in standards and standards reform.* In D. M. McInerney, S. Van Etten, & M. Dowson. *Standards in education* (pp. 3-11). Charlotte, NC: Information Age Publishing.

Ehren, J. S. M. (2015). The evidence base for school inspection frameworks. *Journal of Educational, Cultural and Psychological Studies (ECPS Journal), 1*(12), 43-76.

Field, A. (2009). *Discovering statistics using SPSS* (3rd edition). London: Sage Publication Ltd.

Finn, C. E., Jr., Petrilli, M. J., & Vanourek, G. (1998). *The state of state standards.* Washington, D. C: Thomas B. Fordham Foundation.

Gandal, M. (1996). *Making standards matter, 1996: An annual fifty state report on efforts to raise academic standards.* Washington, DC: American Federation of Teachers.

Goldhaber, D. (2015). *Teacher effectiveness research and the evolution of US teacher policy. The productivity for results series no. 5.* Retrieved from http://files.eric.ed.gov/fulltext/ED560206.pdf

Hamilton, L. S., Stecher, B. M., Yuan, K., Center, E. P., & Corporation, R. (2008). *Standards-based reform in the United States: History, research, and future directions.* Center on Education Policy.

Hammer, D. (1998). *The standards teacher. Standards for excellence in education.* Washington D.C.: Council for Basic Education.

Hanushek, E. A., Peterson, P. E., & Woessmann, L. (2012). *Achievement growth: International and U.S. state trends in student performance.* Retrieved from https://www.hks.harvard.edu/pepg/PDF/Papers/PEPG12-03_CatchingUp.pdf

Harris, A. (2002). *School Improvement: What's in it for school?*. London: Falmer Press.

Harris, A., & Muijs, D. (2005). Improving schools through teacher leadership. London: Open University Press.

Hill, H. C. (2001). Policy is not enough: Language and the interpretation of state standards. *American Educational Research Journal, 38*(2), 289-318.

Jazadi, I. (2003). *An investigation of current constraints and potential resources for developing learner-centred curriculum frameworks for English language at high schools in Lombok, Indonesia.* (Unpublished doctoral thesis). University of South Australia, Adelaide, Australia.

Kaluge, L., & Tjahjono, H. (2004). *The quality improvement of primary children learning through a school-based programme in Indonesia.* A research paper. Universitas Surabaya, East Java.

Kawamoto, K., & Greenes, R. A. (2014). The role of standards: What we can expect and when. In R. A. Greenes (Ed.), *Clinical decision support: The road ahead* (2nd ed., pp. 599–615). Elsevier. Retrived from http://dx.doi.org/10.1016/B978-0-12-398476-0.00021-X.

Kompas. (2010, May). *Foke prihatin, 39.179 siswa SMP gagal UN.* Retrieved from http://edukasi.kompas.com/read/2010/05/06/16552537/

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103-152.

Kyriakides, L., Campbell, R. J., & Gagatsis, A. (2000). The significance of the classroom effect in primary schools: An application of Creemers' comprehensive model of educational effectiveness. *School Effectiveness and School Improvement, 11*(4), 501-529.

Luyten, J. W., & Snijders, T. A. B. (1996). School effects and teacher effects in Dutch elementary education. *Educational Research and Evaluation, 2*(1), 1-24.

Marzano, R. J. (2007). *The art and science of teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., & Kendall, J. S. (1996). *The fall and rise of standards-based education. Issues in brief.* Alexandria, VA: National Association of State Boards of Education.

Maulana, R., Opdenakker, M., den Brok, P., & Bosker, R. (2011). Teacher-student interpersonal relationships in Indonesia: Profiles and importance to student motivation. *Asia Pacific Journal of Education, 31*(1), 33-49.

McClure, P. (2005). Where standards come from. *Theory into Practice, 44*(1), 4-10.

Mohandas, R. (2004). *Overview of education for sustainable development (ESD) in Indonesia.* Jakarta: National Office for Educational Research and development, The Ministry of National Education.

Muijs, D., & Reynolds, D. (2000). School effectiveness and teacher effectiveness in Mathematics: Some preliminary findings from the evaluation of the Mathematics Enhancement Program (Primary). *School Effectiveness and School Improvement, 11*(3), 273-303.

Neumann, K., Fischer, H. E., & Kauertz, A. (2010). From PISA to educational standards: The impact of large-scale assessments on science education in Germany. *International Journal of Science and Mathematics Education, 8*(3), 545-563.

NRC. (2001). *National science education standards.* National Research Council. Washington DC: National Academy Press.

Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes, 18*(2), 161-175.

PI. (2005). *Peraturan pemerintah republik Indonesia Nomor 19 Tahun 2005 Tentang Standar Pendidikan Nasional.* Pemerintah Indonesia. Jakarta.

Priyanto, A. D. (2009). Maximizing SFL contribution to ELT Indonesia. *UAD TEFL Conference.* Yogyakarta: UAD Press.

Raizen, S. A. (1998). Standards for science education. *Teachers College Record, 100*(1), 66-121.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). *MLwiN Version 2.0.* Centre for Multilevel Modelling, University of Bristol.

Ravitch, D. (1995). *National standards in American education: A citizen's guide.* Washington, DC: Brookings Institution.

Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*(2), 197-230.

Scheerens, J. (2013). The use of theory in school effectiveness research revisited. *School Effectiveness and School Improvement, 24*(1), 1-38.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon etc

Schmidt, W. H., Houang, R., Shakrani, S., & Fordham, T. B. (2009). *International lessons about national standards.* Thomas B. Fordham Institute.

Schmoker, M., & Marzano, R. J. (1999). Standards-based education. *Educational Leadership, 56*, 17-21.

Setiogi, S. P. (2003, December 6). Final exam results show real educational quality. *The Jakarta Post*. Retrieved from https://www.thejakartapost.com/news/2003/06/12/final-exam-results-show-real-education-quality.html

Stosich, E. L. (2016). Building teacher and school capacity to teach to ambitious standards in high-poverty schools. *Teaching and Teacher Education, 58,* 43-53.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research.* London: Falmer Press.

Tilaar, H. A. R. (1992). *Manajemen Pendidikan Nasional, kajian pendidikan masa depan* Bandung: Remaja Rosdakarya.

Utomo, E. (2005). *Challenges of curriculum reform in the context of decentralization: The response of teachers to competency based curriculum (CBC) and its implementation in schools.* (Unpublished doctoral dissertation). University of Pittsburgh, USA.

Van der Werf, M. P. C., Creemers, B. P. M., De Jong, R., & Klaver, L. (2000). Evaluation of school improvement through an educational effectiveness model: The case of Indonesia's PEQIP project. *Comparative Education Review, 44*(3), 329–356.

Volante, L. (Ed.). (2012). *School leadership in the context of standards-based reform: International Perspectives.* New York, NY: Springer Science and Business Media.