Technology, Pedagogy and Education

ISSN: 1475-939X (Print) 1747-5139 (Online) Journal homepage: https://www.tandfonline.com/loi/rtpe20

Does smartphone-assisted student feedback affect teachers' teaching quality?

Hannah J.E. Bijlsma, Adrie J. Visscher, Marjoleine J. Dobbelaer & Bernard P. Veldkamp

To cite this article: Hannah J.E. Bijlsma, Adrie J. Visscher, Marjoleine J. Dobbelaer & Bernard P. Veldkamp (2019): Does smartphone-assisted student feedback affect teachers' teaching quality?, Technology, Pedagogy and Education, DOI: 10.1080/1475939X.2019.1572534

To link to this article: https://doi.org/10.1080/1475939X.2019.1572534

Published online: 27 Feb 2019.

TECHNOLOGY, PEDAGOGY AND EDUCATION
https://doi.org/10.1080/1475939X.2019.1572534

Routledge
Taylor & Francis Group

Check for updates

Does smartphone-assisted student feedback affect teachers' teaching quality?

Hannah J.E. Bijlsma[a], Adrie J. Visscher[a], Marjoleine J. Dobbelaer[a] and Bernard P. Veldkamp[b]

[a]Department of Teacher Development ELAN, University of Twente, Enschede, the Netherlands ; [b]Department of Research Methodology, Measurement, and Data Analysis (OMD), University of Twente, Entschede, The Netherlands

ABSTRACT

In this study, it was investigated if student feedback promoted teachers' insight into where they could improve their lessons and their reflection on their lessons. It was also studied in what ways teachers worked on improvement, based on the student feedback, and whether it affected the quality of their teaching. Student perceptions of teaching quality were measured by means of a smartphone application for providing teachers with feedback on their lesson. Teachers in this study (N = 60) were randomly assigned to the control or experimental group. The smartphone application was used several times by teachers to obtain student feedback. The teachers reported that they gained insight into where they could improve their lessons. They reported improvement-oriented actions in response to the student feedback. According to students, teachers first slightly improved their teaching quality. However, teachers did not seem to reflect significantly more on their lessons and their improvement did not sustain. Explanations for the findings are discussed and suggestions for future research are presented.

ARTICLE HISTORY
Received 1 February 2018
Accepted 12 October 2018

KEYWORDS
Digital feedback system; student feedback; student perceptions; teaching quality

Introduction

Research shows that feedback can have a strong performance-enhancing effect (Hattie & Timperley, 2007; Ilgen, Fisher, & Taylor, 1979; Kluger & DeNisi, 1996). In education, however, although teachers do receive immediate feedback during the lessons, teachers do not obtain much structured feedback on their functioning (Sawada et al., 2002; Sluijsmans & Kneyber, 2016; Voerman, Meijer, Korthagen, & Simons, 2012). One way for teachers to obtain feedback is through lesson observations (Lasagabaster & Sierrra, 2011), but reliable and valid observation scores about teaching quality require multiple observations and multiple, trained raters (Hill, Charalambous, & Kraft, 2012; Muijs, 2006). This makes lesson observations expensive and time-consuming.

Another, more efficient (less time-consuming, less expensive and faster) way to provide teachers with feedback is by measuring student perceptions of the teaching they have experienced during the lesson(s) of their teachers (Muijs, 2006; Peterson, Wahlquist, & Bone, 2000). If student perceptions are used, the number of observations (in cases where students evaluate the teaching of the teacher across many lessons) and observers (the number of students) is larger than in the case of lesson observations, which could improve the reliability of the scores (Fauth, Decristan, Rieser, Klieme, & Büttner, 2014). In addition, student perceptions reflect the perspective of the target group (Kane & Staiger, 2012; Quaglia & Corso, 2017) and thereby promote *student voice* (Cook-Sather, 2002, 2007): the voice of students in their education (Lincoln, 1995). Teachers can use this feedback to improve their teaching quality.

CONTACT Hannah J.E. Bijlsma h.j.e.bijlsma@utwente.nl

Despite these advantages of using student perceptions for obtaining feedback, several concerns are at stake. Students are not fully objective raters because they are closely involved in the lessons and they have their personal, subjective preferences (MacNell, Driscoll, & Hunt, 2015). Teachers' scores may be biased because of, e.g., the popularity of the teacher, or students' grades for a subject (Ferguson, 2012; Van der Lans, Van de Grift, & Van Veen, 2016). Other concerns relate to discriminant validity (students' ability to distinguish between different teaching dimensions) and the extent to which students' ratings of a teacher correlate with teacher ratings by external observers (Van der Lans, 2017). Although several studies show that student perceptions can provide reliable (Fauth et al., 2014; Kane & Staiger, 2012) and valid (Balch, 2012; Peterson et al., 2000; Van der Scheer, Bijlsma, & Glas, in press) information about teachers' teaching, some publications are more critical on the use of student perceptions (Cantrell & Kane, 2013; Roche & Marsh, 2000).

Technological developments make it possible for teachers to obtain feedback from students in an immediate and inexpensive way by means of a smartphone application. This way, teachers can obtain feedback about one single lesson. Because the feedback is provided to the teacher right after the lesson, the link between the actual acting of the teacher in the classroom and the student feedback is clearer than in the case of feedback on teacher behaviour in general (across many lessons; Hattie & Timperley, 2007; Shute, 2008). When feedback is given immediately, it is found to be more effective than when it is postponed (Timmers & Veldkamp, 2011). Teachers might therefore be able to work better on improving their teaching quality.

To our knowledge so far, little research has been done on the effects of using structured student feedback on teaching quality, especially when it is obtained by teachers in an immediate, direct way, and linked to a specific lesson (instead of about a teacher's general teaching quality). In the current study, students' perceptions of one lesson were measured and the feedback was presented to teachers to inform them on the strengths and weaknesses of their lessons. It was hoped that the feedback would promote teachers' professional reflection on their teaching quality (reflection does not happen spontaneously; Driessen, Van Tartwijk, & Dornan, 2008; Ertmer & Newby, 1996) and that teachers would attempt to improve their teaching. The following research questions were formulated:

Does smartphone-assisted student feedback promote:

- teachers' insight into where their lessons can be improved,
- their professional reflection,
- improvement-oriented actions, and
- does it affect the quality of their teaching?

Theoretical framework

Student feedback for improving teaching

Based on more than a hundred years of feedback research, we know that feedback can have a strong performance-enhancing effect (Kluger & DeNisi, 1996). Feedback is conceptualised as information provided by an agent (e.g. teacher, student, peer, self, experience) regarding aspects of one's performance or understanding (Hattie & Timperley, 2007). The feedback a recipient receives can support changing his/her ways of acting and improve his/her performance.

As it was stated above, teachers do not receive much feedback on their professional functioning, and therefore, student perceptions of teaching quality can function as a valuable source of feedback for teachers (Muijs, 2006; Peterson et al., 2000). But in what way can student feedback help improve teachers' teaching quality? This is explained by Fraser (2007), who studied the use of student feedback to teachers for improving their teaching. The feedback in this case was about teachers' contribution to a positive and safe learning environment in the classroom and about the nature of teachers' teaching activities during the lessons (e.g. how they provided instruction, what kind of questions they asked their students and in what way). Fraser (2007) argued that the improvement process based on student feedback includes five steps:

(1) Data on student perceptions of the lesson are collected.
(2) The results are provided to teachers as feedback.
(3) Teachers can identify aspects of their lessons which need improvement and consider alternative ways of acting. In other words: the feedback can stimulate teachers to reflect on their own practices.
(4) Teachers can carry out improvement-oriented actions during and outside the lessons over a period of time, based on the feedback.
(5) To determine the effectiveness of the actions undertaken by teachers, student feedback can be collected again.

Thus, by making teachers reflect on their own practices (Driessen et al., 2008; Ertmer & Newby, 1996), the feedback can provide an alternative perspective on teachers' teaching (Bell & Aldridge, 2014; Hoban, 2004; Hoban & Hastings, 2006). Teachers can develop improvement-oriented actions as a follow-up to their professional reflection on their teaching (Fraser's fourth step). This can be actions *during the lesson*, in which a teacher directly tries to improve something. For example, giving students more time in the lesson to work on assignments, or adapting the pace of the lesson to students' needs. Student feedback can also lead to improvement-oriented actions *outside the lessons*, e.g. when teachers consult colleagues for advice, attend professionalisation activities or look for relevant information on the Internet or elsewhere. Teachers can also discuss the feedback with their students (as a group or individually; Gaertner, 2014). This process can make teachers improve their teaching quality (visualised in Figure 1).

Figure 1. Theory of action.

Factors influencing feedback

So far, we have described a theory of action for the improvement of teaching quality when receiving student feedback. However, we know that the effectiveness of feedback is not straightforward but depends on several factors (Hattie & Timperley, 2007; Visscher, 2015). The timing, frequency and content of the feedback are examples of influencing factors: when (immediately or postponed) and how often is the feedback provided to teachers, and what information does the feedback contain (Shute, 2008)? According to Timmers and Veldkamp (2011), feedback is preferably given on regular times, right after the specific lesson, so the link between the lesson and the feedback can be made clearer than when the feedback is postponed. The feedback ideally does not only inform the level of performance but also, in case there is room for improvement, how the recipient can improve.

Contextual factors also play an important role in teachers' use of the feedback and the choices that they make regarding improving teaching (Levin & Datnow, 2012; Schildkamp & Lai, 2013; Wohlstetter, Datnow, & Park, 2008). School leaders, for example can facilitate and support teachers to work collaboratively on team-shared goals, or on goals related to school-specific policies, and they can create a culture of teacher professionalisation.

Recipient characteristics also influence the use of feedback. Teachers can accept or reject the student feedback as information about their teaching quality, and they need to be motivated to utilise the data (Fresko & Nasser, 2001; Gaertner, 2014; Lincoln, 1995; Miller, 1971; Pambookian, 1972, 1976; Tuckman & Olivier, 1968). Also, teachers' background characteristics influence the use of feedback by teachers (e.g. their teaching quality and teaching experience, their gender and age). These recipient characteristics are in line with general theory of the persuasive system design (Oinas-Kukkonen & Harjumaa, 2009). In the Fogg Behavioural Model (Fogg, 2009) it is argued that 'for a person to perform a new behaviour, he or she must (1) be sufficiently motivated, (2) have the ability to perform the behaviour, and (3) be triggered to perform the behaviour' (Fogg, 2009, p. 1).

Hypotheses

Based on the improvement process as described by Fraser (2007), and the chain of events presented in Figure 1, we were now especially interested in the effects of using smartphone-assisted student feedback on teachers' insights into areas for improvement, reflection on their lessons, teachers' improvement-oriented actions, and on the quality of their lessons. We decided to provide teachers with the feedback without any further support and not taking any of the influencing factors into account yet (this will be done in other papers). The following four hypotheses were tested (the article is organised in line with these hypotheses).

The provision of smartphone-assisted student feedback:

(1) … enhances teachers' insight into where their lessons can be improved;
(2) … enhances teachers' professional reflection on their lessons;
(3) … increases teachers' improvement-oriented actions;
(4) … improves teachers' teaching quality.

Method

The Impact! tool

Feedback was provided to teachers by means of the Impact! tool, a smartphone application which can be used on a student device (smartphone, iPad or tablet). Students sign in on the application (or web app) and teachers have access to a private web environment. Teachers choose the lesson they want to receive feedback on and in that lesson, students are asked to answer the items in the tool. The items measure how a student perceives the lesson, for example whether students think that the lesson connected well with their prior knowledge. Figure 2 shows an example of an Impact! item on a student phone (in the study, questions were presented in Dutch). Students answer the questions on a 4-point Likert scale (*totally agree*, *agree*, *disagree* or *totally disagree*). If a question does not apply, the option 'not applicable' (Niet van Toepassing in Dutch) can be used (this option was added to three questions). The 'i' and '?' buttons provide additional information about how to use the tool.

Every time the tool has been used, teachers receive a summary of the feedback the students have given (for examples, see figures 3a and 3b). The feedback is confidential: the teacher does not know what an individual student answered. The web environment makes it possible for teachers to organise students into groups, based on students' performance levels for the subject taught by the teacher (high, middle and low level). As a result, students' responses on an item are displayed per student performance group (see Figure 3a; different colours represent different groups). If the tool is used multiple times, then the development in the scores can be displayed per item (Figure 3b).

Development of the application and the content of the questionnaire

A company developed the technical part of the Impact! tool, based on the specifications as made by the authors of this article. The technology was tested in different classrooms until it worked perfectly.

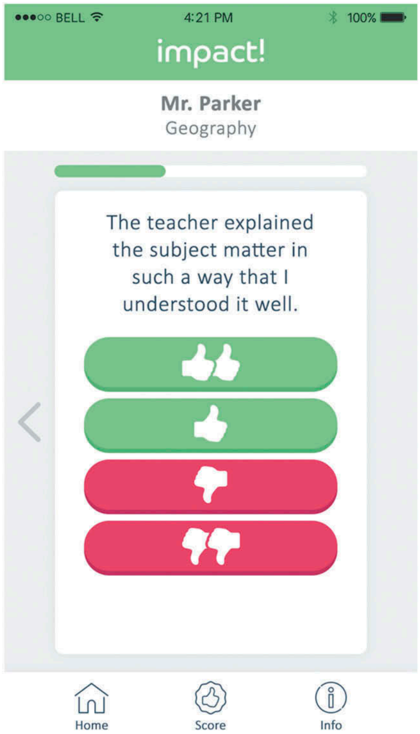
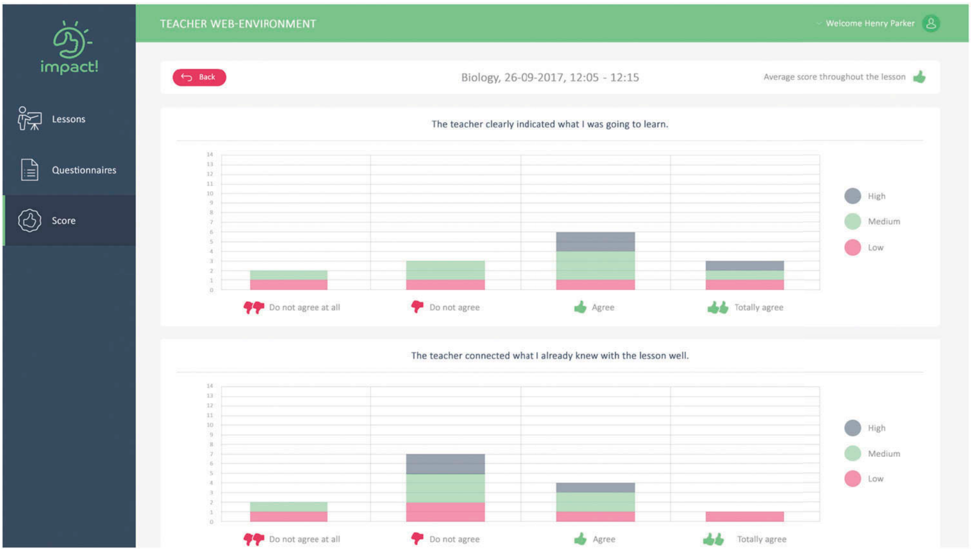


Figure 2. Example of an item on a student phone screen.

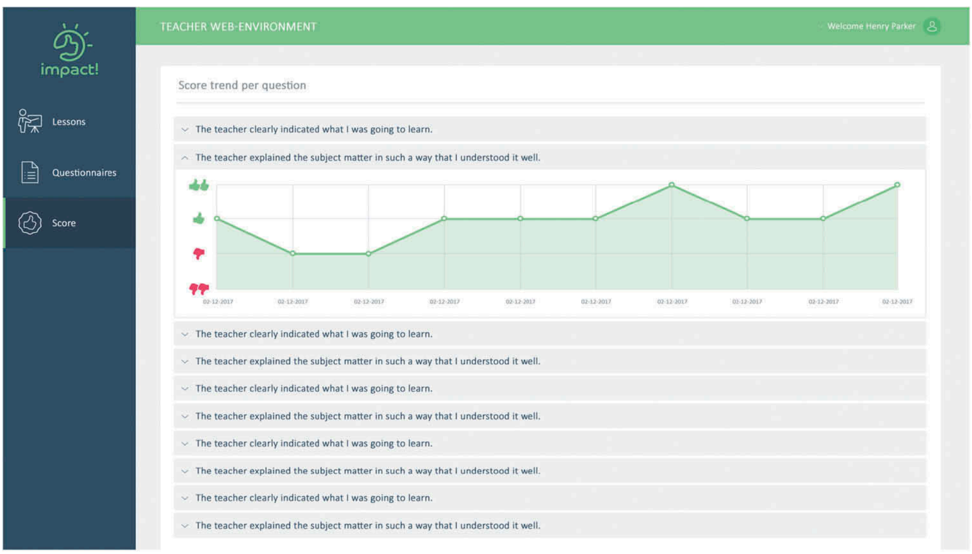
For the development of the questionnaire, several considerations were taken into account. To require the least effort from students in answering the items, a standard set of a few items (16) was included in Impact! and one item included only one teaching aspect. Four answer options were given to avoid central tendency in answering the items (Weisberg, 1992). Students were asked to rate a lesson on scientifically proven characteristics of effective lessons (instead of, for example, asking them if they like the lesson) and all items were about one single lesson (instead of about the lessons of a teacher in general). They were all formulated in a teacher-centred way (e.g. '*The teacher* created a safe atmosphere during the lesson') to determine the contribution of the teacher to lesson quality. Because students were supposed to give their own opinion when answering the items, 'I' was used instead of 'our class' (e.g. 'The teacher clearly indicated what *I* was going to learn'). To enable teachers to ask more than only the standard set of items, they had the opportunity to add questions to the standard set of items. To present students from becoming bored and not reading the items properly when filling out the questionnaire, the order of the items changes every time the tool is used.

For the development of the content of the items of the Impact! tool, first, the literature about effective teaching was studied. The results of several meta-analyses about effective teaching show a variety of teaching practices which are known to be effective for student learning (e.g. Creemers, 1994; Fauth et al., 2014; Muijs et al., 2014; Pianta & Hamre, 2006; Reynolds et al., 2014; Sammons, Hillman, & Mortimore, 1995). For this study, these practices were categorised into seven general teaching dimensions (Day, Sammons, & Kington, 2008; Marzano, 2003):

(1) creating a supportive and positive classroom climate (Fraser, 1998);
(2) having a well-organised and structured classroom management (Day et al., 2008; Marzano, 2003; Pianta & Hamre, 2006);



a) A teacher's app scores for specific items; students' levels of performance are displayed in grey (high), green (middle) and pink (low).



b) Trend overview of how a teacher performs over time on one specific item (class average scores).

Figure 3a and 3b. Examples of the feedback in the teacher web environment.

(3) providing clear instruction (Hollingsworth & Ybarra, 2015; Muijs et al., 2014; Rosenshine, 1995);
(4) adaptive instruction (Maulana, Helms-Lorenz, & Van de Grift, 2015; Van de Grift, 2007);
(5) the quality of the teacher–student interaction (Den Brok, Brekelmans, & Wubbels, 2007);
(6) cognitive activation of students to promote deep learning (Drijvers, 2015; Hiebert & Grouws, 2007);
(7) assessing students' learning during the lesson (formative assessment; Black & Wiliam, 2006; Hattie, 2008; Hattie & Timperley, 2007).

A review of student perception questionnaires was also conducted (Bijlsma, 2016) to gather information about items used in other questionnaires. Both, the literature study and the review, led to ideas about and formulations of items for the Impact! tool. Potential items were discussed extensively among the authors. Experts in the field of educational sciences contributed to our study by identifying five items which in their opinion should be included in the questionnaire. Moreover, teachers and students from Dutch secondary schools (Grade 9) were asked whether the items were comprehensible and clear, and what they thought the items were about. Based on their feedback, some items were reformulated which resulted in a final set of 16 items (see Table 1).

Research design and intervention

An experimental design was chosen for this study as this was considered to be the best research design for evaluating the effects of smartphone-assisted student feedback (Fraenkel, Hyun, & Wallen, 2012). From some schools, more than one teacher participated in the study. To avoid cross-over effects (teachers of the same school influencing each other; Fraenkel et al., 2012),

Table 1. Teaching aspects (items) related to the teaching dimensions and practices.

	Teaching aspects (items)	Teaching dimension	Related teaching practices
1	The teacher clearly indicated what I was going to learn.	Clear instructional approach.	Share learning objectives with students at the beginning of the lesson; have a clear beginning of the lesson.
2	The teacher explained the subject matter in such a way that I understood it well.	Clear instructional approach.	Introduce new material in small steps; give examples and model procedures; use clear language; ensure guided practice with students.
3	The teacher connected what I already knew with the lesson well.	Clear instructional approach.	Review on, and connect to, prior knowledge of students.
5	The teacher asked questions about the subject matter, which made me think.	Cognitive activation and deep learning.	Ask high-qualitative questions frequently to students in order to promote deep learning; model assignments; solve subject-matter-related problems together with students.
6	If I did not understand the subject matter, the teacher made sure I understood it.	Adaptive instruction.	Differentiate instruction; processing and/or assessment and adapt activities to the varying needs of students.
7	The teacher made sure we worked hard during the lesson.	Well-organised and structured classroom management.	Ensure an efficient classroom management and organisation; correct disruptive behaviour effectively; have clear classroom rules and routines.
8	The teacher made sure I was not afraid to say if I did not understand something.	Supportive and positive classroom climate.	Create a safe learning environment during the lesson.
9	The teacher created a good classroom climate.	Good student–teacher interaction.	Teachers are supporting and leading.
10	The pace of the lesson was good for me.	Adaptive instruction.	Adapt the pace of the lesson to the different learning needs of students.
11	If I gave an incorrect answer, the teacher explained why it was wrong.	Clear instructional approach.	Correct students' incorrect answers; repeat instruction (in a different way, when necessary).
12	The teacher ensured my active participation in the lesson.	Clear instructional approach.	Make use of activating learning activities during the lesson.
13	The teacher gave enough time to work on the assignments in the lesson.	Well-organised and structured classroom management.	Provide extended practice and transfer by assignments and ensure independent practice of students.
14	At the end of the lesson, the teacher summarised what we had learned.	Clear instructional approach.	Have a clear ending of the lesson; review on the goal of the lesson with students and check whether the goal is achieved by most students.
15	I now can complete assignments about the subject matter on my own.	This item is meant to measure the effect of the lesson (did the lesson positively affect student learning?).	
16	The teachers' lessons will become even better if he/she…	This is an open question. Students can type their answer in the app to give a tip to the teacher about improvement of the lesson.	

schools were randomly assigned to the experimental group or the control group. Therefore, the number of teachers in both groups was slightly different.

During the intervention period of four months, teachers in the experimental group used the Impact! tool and thus received feedback on their lessons from their students. Teachers in the control group did not use the Impact! tool. They were given access to the tool after the intervention and the post-test.

The project timeline is presented in Figure 4. M0, M1, M2 and M3 indicate the measurement moments, where M0 is the pre-test, M3 is the post-test, and M1 and M2 are in-between measurements. The green line indicates the use of the Impact! tool and the green stars indicate the lessons in which the tool was used. The number of times the tool was used differed per teacher, because it was up to the teacher to decide when to use it. Therefore, the number of stars varies between teachers. For example, the five first stars in the figure can be interpreted as one teacher who used the tool five times between M0 and M3.

Figure 4. Measurement moments in the experimental group and the control group.

Participants

In total, 58 mathematic teachers participated in the study. Mathematics was chosen as a subject for this intervention study because in general, math lessons have a clear structure and a clear moment of instruction. Mathematics is an obligatory school subject for all students. It is regarded as one of the most important subjects in secondary schools (in addition to reading, writing and spelling).

Regular Dutch secondary schools were contacted by email and phone and teachers were also contacted individually, inviting them to participate in the study. Teachers were informed about the study design, about what was expected from them, and about the Impact! tool.

Twenty-eight teachers from 23 schools were assigned to the experimental group and 32 teachers from 25 schools were assigned to the control group. One teacher in the experimental group stopped for health reasons. One other teacher was excluded from the analyses as this teacher did not fill out the questionnaires at all despite multiple requests to do so. On both, M2 and M3, 24 teachers from the experimental group filled out the digital questionnaire.

An overview of the background characteristics of the 60 teachers in both groups is presented in Table 2. On M2, M3 and M4, respectively 58%, 58% and 60% were male, and the other characteristics remain stable. It shows that both groups are similar with respect to teachers' gender, age and nationality. Teachers in the control group had slightly more years of general teaching experience, but this difference is not statistically significant (F = 1.457; p = .232).

A total of 1488 students participated in the study (14/15 years old; 48.5% boys in M0), with 681 students assigned to the experimental group and 807 students assigned to the control group. Students were introduced into the use of the Impact! tool. To encourage them to answer the Impact! questions from their perspective, no explanation was given about the content of the questions.

Table 2. Background characteristics of the 60 teachers (on M0).

		Experimental group			Control group		
		N	(%)	*M (SD)*	*N*	(%)	*M (SD)*
Schools		23	(47.92)		25	(52.08)	
Teachers		28	(46.77)		32	(53.33)	
Teacher gender	Female	12	(42.90)		13	(40.60)	
	Male	16	(57.10)		19	(59.40)	
Teacher age				41.21 (2.05)			41.00 (1.93)
Years of teaching experience	General			11.23 (1.47)			14.00 (1.71)
	Math (*SE*)			10.84 (1.51)			11.09 (1.46)
Nationality	Dutch	31	(96.90)		28	(100.0)	
	Other	1	(3.10)		0	(0.00)	

Instruments

Questionnaires were administered to teachers and students in both groups for the pre-test and post-test (M0 and M3). These questionnaires included, first of all, questions on teacher background characteristics.

To measure *professional reflection* by teachers, the four-item reflection scale of the Level of Reflective Thinking questionnaire was administered (Kember et al., 2010). For the purpose of this study, the Dutch translation of this questionnaire by Könings et al. (2016) was adapted for use in educational settings (e.g. 'I often reflect on my actions *in the lesson* to see whether I could have improved on what I did'). The items were answered on a 4-point Likert-scale. The factor structure of the full questionnaire as described by Kember et al. (2010) was considered to be good. In line with Bron, Endedijk, Van Veelen, and Veldkamp (2018), a confirmative factor analysis was conducted with the four newly formulated items, to verify that the Impact! items form one scale (which makes it useful to report a sum score). The results showed that the items loaded fairly well on one scale (factor loadings ranged from .61 to .83). The reliability of the scale for the scores on M0 was reasonable (Cronbach's alpha = .71).

To measure teachers' *teaching quality*, students completed a questionnaire on M0 and M3. In the Impact! tool, the questions are about one specific lesson, whereas in the questionnaires used on M0 and M3, the same items were about teachers' general teaching practices. Examples of items, answered by students on a 4-point Likert-scale, are: 'The teacher clearly indicates in the lessons what I am going to learn' and 'The teacher usually makes sure we work hard during lessons.' An explanatory factor analysis (Bron et al., 2018) showed that all items loaded moderately to well on one scale (factor loadings ranged from .43 and .77). Reliability was considered good (α = .89).

In both of the questionnaires that were just mentioned, reliability was investigated by calculating Cronbach's alpha. Although Cronbach's alpha receives quite some critique (e.g. it might underestimate reliability; Dunn, Baguley, & Brunsden, 2013, Sijtsma, 2009), it is widely used as a measure for reliability (COTAN, 2010). It reflects the internal consistency or average correlation between the items in a survey instrument (Santos, 1999). Moreover, according to De Vocht (2016), Cronbach's alpha can be used reliably if the items all relate to a single concept (i.e. they are unidimensional), if the coding of the items are all the same and if the sum score measures one underlying construct. These conditions are met in this study.

Teachers in the experimental group completed two additional, digital questionnaires during the intervention period about the use of the tool (M1 and M2). In these digital questionnaires, and in the questionnaire on M3, teachers in the experimental group reported whether they had obtained *insight into where they could improve* their lessons based on the tool feedback. In addition to this, teachers reported whether they had undertaken *improvement-oriented actions* in response to the feedback during the follow-up lessons (e.g. 'At the end of the lesson, summarise what students had learned', in response to item 14 of the Impact! tool) and/or outside the lesson (e.g. 'I discussed the feedback with the students', or 'I discussed the feedback with my supervisor').

Data analysis

Descriptive statistics were calculated for the frequency of use of the Impact! tool by teachers in the experimental group. A minimum of four and a maximum of sixteen measurements (tool use frequencies on M0 and M3) per teacher were included, because only those measurements in which at least one-third of the students gave feedback about teachers' teaching quality were used in the analysis (calculated by the number of tool results plus the additional paper-based student feedback, compared with the total number of students in the class). Students who were not attending the lesson were not notified to fill in the questionnaire.

To test hypothesis one of this study, bar charts were made for *teachers' insight into where they can improve* their lesson (as reported by teachers) and the results are presented. The development of *teachers' professional reflection* (hypothesis two) was analysed by conducting an analysis of covariance (ANCOVA), using SPSS Statistics version 24. For the variable intensity of reflection (on M3 for all participating teachers), the assumptions for ANCOVA were checked. Independent measurements were guaranteed by the sampling procedure (randomised control). By looking at the Q–Q plots, the variable seemed to have a normal distribution. Levene's test showed an equality of variances (F = .191; p = .664). It is assumed that covariates have been measured without errors. In the ANCOVA, participation in the intervention was included as a factor. The means on M3 for intensity of reflection in both the experimental group and the control group were used for measuring the dependent variable. The means on M0 for intensity of reflection were included as a covariate. Bar charts were made to examine the number of improvement-oriented actions undertaken by teachers in the experimental group *during the lesson* (hypothesis three) and the results were presented. To examine the relation between teachers' initial teaching quality and the improvement-oriented actions teachers undertook, scatterplots were made. Only teachers who completed the digital questionnaires on M1, M2 and M3 were included in this analysis (N = 20). Also, teachers' improvement-oriented actions undertaken *outside the lesson* were examined by means of bar charts and presented in the results section.

To test hypothesis four, graphs of *teachers' teaching quality* over time (across all teaching aspects as well as for every teaching aspect separately) were made and studied. The graphed scores are averages of teachers' teaching quality as reported by students over a five-day period. That way, the scores for one working week were compressed in order to reduce data and to enable a better examination of the trends. Scores on the students' questionnaires about their teachers' teaching quality on M0 (pre-test) and M3 (post-test) were included, as well as all scores from the Impact! tool. The number of measurements per time period of five days was, on average, 7 (SD = 3.92).

A multilevel growth model (Field, 2013; Snijders & Bosker, 1999) was made as it takes into account the nested structure of the data (different time points were nested within one teacher; Peugh, 2010). Consistent with previous research (Leatherdale, Cameron, Brown, & McDonald, 2005; Van der Scheer & Visscher, 2017), a four-step modelling procedure was conducted as a follow-up to the null model, using SPSS Statistics version 24. Scores for teaching quality on M3 were used as the dependent variables. 'Time' was used as the independent variable (centred at the first measurement, 22 November 2016). Values express the number of days that had elapsed since the first measurement. Linear and quadratic terms of time were included in the models. In the null model, it was determined whether teachers vary in teaching quality scores on M3. Model 1 was developed with only the linear effect of time as a predictor. In model 2, the linear as well as the quadratic effect of time were added as predictors, and predicted scores were saved to be plotted against 'date', and to illustrate the quadratic trend. In models 3 and 4, the strength of the direct effects of both the teacher- and the date-level predictors was computed using random coefficient regression models (to investigate whether teachers have different slopes). Model 3 only included the variable time (centred around the first measurement) as a predictor. For model 4, the variable time was centred quadratically and added as well.

Results

The magnitude of tool use

The number of measurements of teachers' teaching quality varied from 4 to 17 times between teachers with an average of 6.7 times. The mean score of the results of all teachers and across all measurements was 3.19 (*SD* = 0.39).

Hypothesis one: teachers' insight into room for improvement

Figure 5 shows, per teaching aspect on M1, M2 and M3 (the three measurement moments, see Figure 4), teachers' insights into where they saw room for improvement in their teaching.

As can be seen, teaching aspects which refer to the teaching dimensions 'clear instructional approach', 'assessment for learning', and 'cognitive activation and deep learning' were cumulatively mentioned most by teachers as teaching aspects that could be improved (aspects 1, 4, 5 and 14). Insight into the fact that

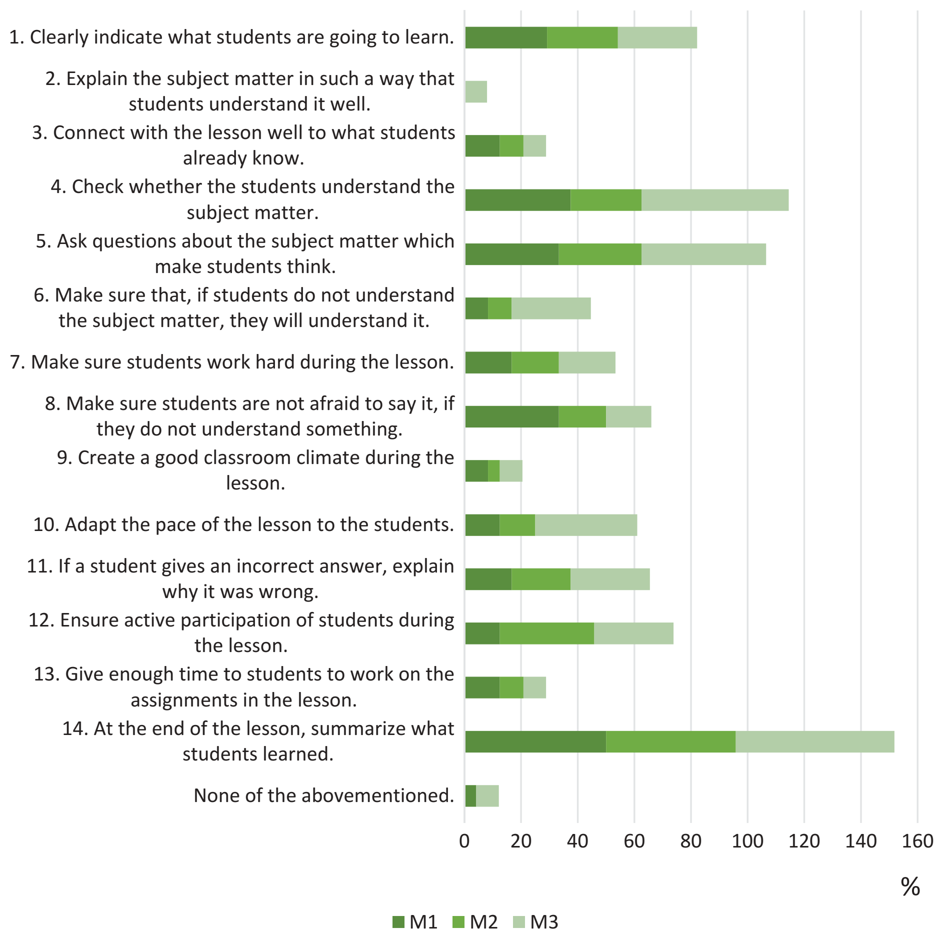


Figure 5. The perceived areas by teachers for improvement (percentage of teachers) on the three measurement moments.

the clarity of the instructional approach could be improved was highest on M3 (as compared to M1 and M2). Some other teaching aspects, related to various teaching dimensions, were not reported much by teachers (e.g. 'explain the subject matter in such a way that students understand it well' and 'create a good classroom climate during the lesson').

Hypothesis two: professional reflection of teachers

The mean score of the control and experimental group was respectively 2.96 (SD = 0.46) and 2.97 (SD = 0.46) on the pre-test and respectively 2.97 (SD = 0.46) and 2.98 (SD = 0.38) on the post-test. The mean difference score between the two groups on M3 was −.00 (SD = 0.41). As regards the development of teachers' reflection on their lesson, no statistically significant difference was found between the experimental group and the control group (F = 0.068; p = 0.796; r^2 = 0.289).

Hypothesis three: teachers' improvement-oriented actions

Improvement actions during the lessons

Figure 6 shows, for M1, M2 and M3, the percentage of the improvement-oriented actions on particular teaching aspects undertaken by teachers during the lessons.

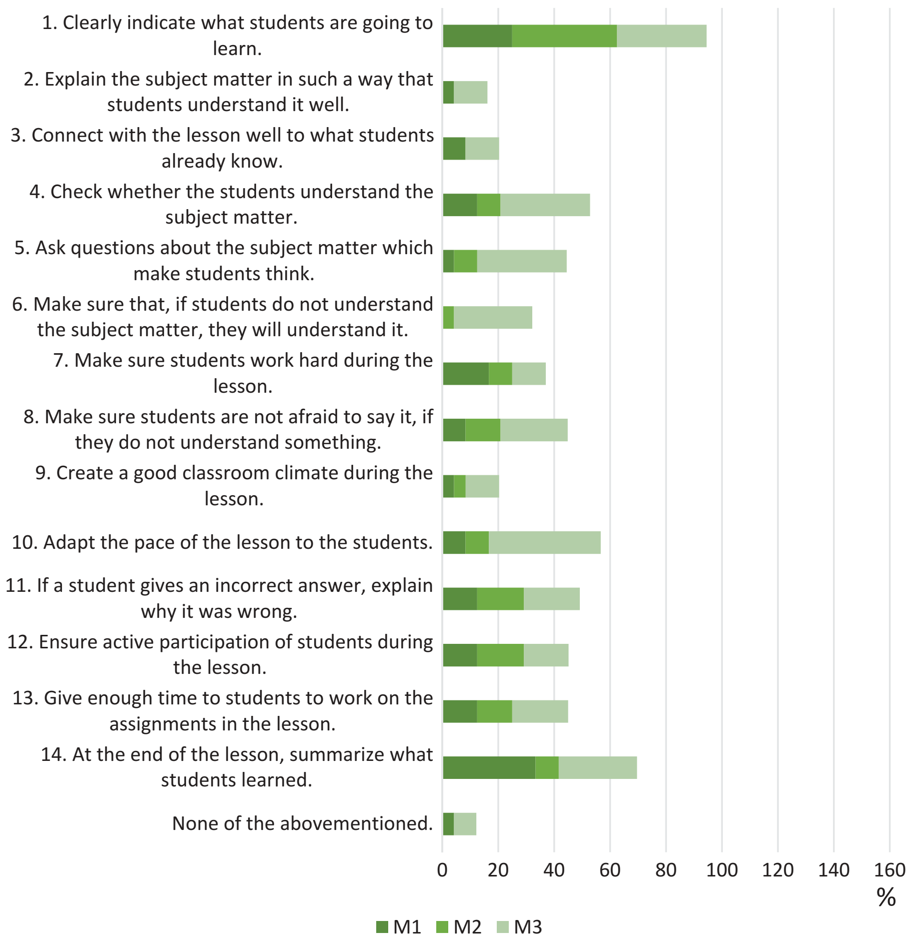


Figure 6. The improvement-oriented actions undertaken by teachers during the lessons at M1, M2 and M3 (percentages of teachers).

Cumulatively, most improvement-oriented actions during the lessons were reported for 'clear instructional approach' (e.g. teaching aspects 1 and 14). Teaching aspects 11, 12 and 13, which relate to 'clear instructional approach' as well as to 'a well-organised and structured classroom management', were reported the same number of times on M1, M2 and M3.

When comparing figures 5 and 6, it can be seen that teaching aspects related to 'good student–teacher interaction' and 'well-organised and structured classroom management' were reported by teachers as teaching aspects where they saw opportunities for improvement *and* they were reported by teachers as teaching aspects they attempted to improve during their lessons (e.g. aspect 9 on M2 and on M3). This was also the case with teaching aspects mainly related to 'clear instructional approach' (e.g. aspects 1, 4, 5 and 14).

Improvement actions outside the lesson

Figure 7 shows the numbers of improvement-oriented actions undertaken by teachers outside the lesson for M1, M2 and M3.

As can be seen, a large percentage of teachers looked at the feedback on M1, M2 and M3. They talked with students about the tool results more frequently as the intervention period progressed (respectively 25%, 28% and 44% of the teachers). A relatively small percentage of teachers discussed the feedback with their supervisor or with other colleagues. The number of teachers that reported that they talked with their colleagues about the feedback was more on M3 (24%) than on M1 (16%) and M2 (12%). No teacher asked colleagues for tips on how to improve, or signed up for a course. The percentage of teachers who looked on the Internet for information about how to improve declined somewhat during the intervention period (from 12% on M1 to 4% on M3). A small percentage of teachers reported that they had not done anything with the feedback (respectively 8%, 4% and 8% of the teachers). Other actions reported by teachers were about using the feedback for formulating learning goals (for internships) and for personal reflection on their lessons.

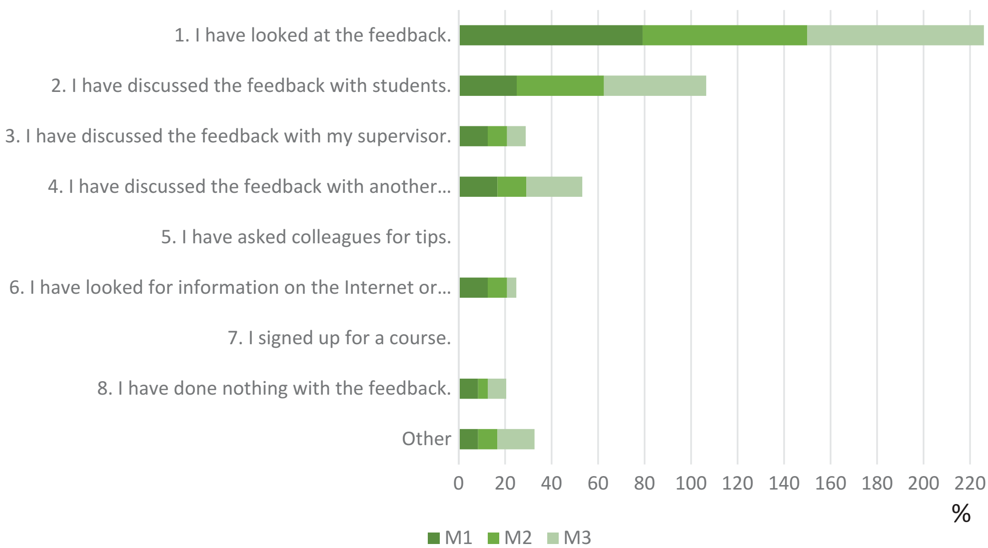


Figure 7. Percentage of teachers who reported they undertook the improvement-oriented actions outside the lesson on the three measurement moments (M1, M2 and M3).

Hypothesis four: improvement of teaching quality

Figure 8 shows teachers' average teaching qualities.

In Figure 8, we see that the average teaching quality improved at the beginning of the intervention period. Just before period 7, teaching quality shows a slight decline. It improved again between period 10 and period 17. After that, teaching quality showed a capricious trend and dropped to about the initial level. Studying the different teaching aspects separately, this trend is also seen for teaching aspects related to the teaching dimension 'clear instructional approach'. According to students, teachers improved on this teaching dimension at the beginning of the intervention period, however the improvement is not sustainable. Teaching aspects related to the teaching dimension 'supportive and positive classroom climate' showed a slight decline at the beginning of the intervention period.

Table 3 shows the results of five multilevel models (including the null model) to estimate teaching quality when taking time into account.

In the null model, we see that teachers vary considerably in teaching quality. Most of the variance (73.4%) in the measurements is situated at the teacher level (intercept variance). Thus, most of the variance is teacher specific and only a relatively small part relates to the residual variance. In model 1, no significant effect was found for time (in linear terms). In model 2, the quadratic term of time was added, which showed a significant effect of this parameter ($p < 0.01$). The variance of 'time' declined from .026 to .020, so the quadratic trend investigated with this model partly explained the variance. Models 3 and 4 were estimated to investigate the existence of random slopes. Because both models show no statistically significant effects, there is no evidence for such random slopes, which implies that teachers do not have different growth curves. The results in Table 5 show a best fit for model 2, because the deviance statistic is lowest there, and parameters were parsimoniously added in the model. The predicted values of teachers' teaching quality of model 2 were plotted against date, which shows a quadratic trend of the predicted teaching quality over time (see Figure 9) and therefore supports the results as presented in Figure 8.

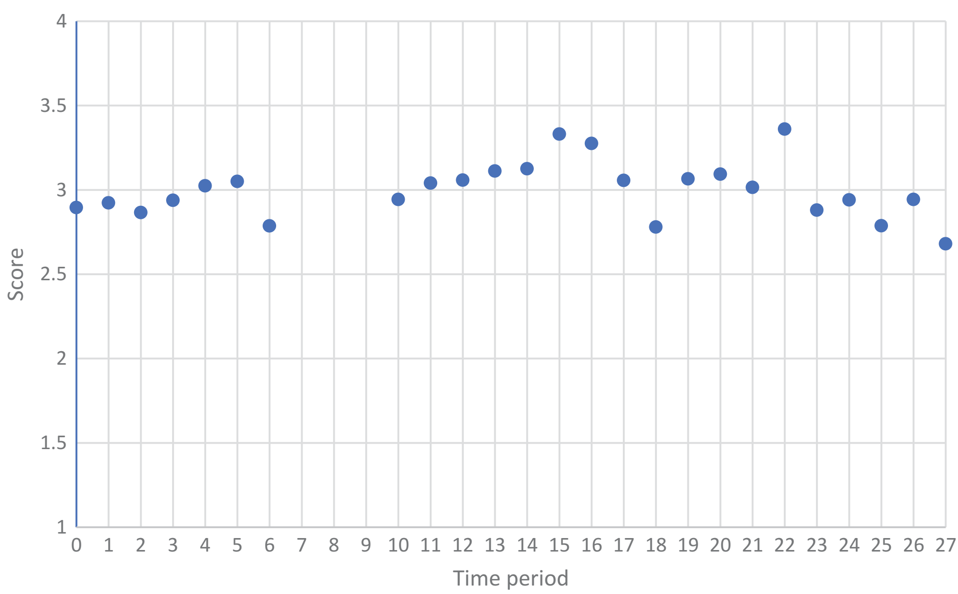


Figure 8. Average teaching quality (a time period includes five days). Due to a holiday period in the Netherlands, no measurement was conducted in time periods 7, 8 and 9. Time period 18 was a holiday period as well, but because not all schools are off in the same week, two measurements were conducted in that period explaining the small decline in this period.

Table 3. Multilevel models.

	Null model	Model 1	Model 2	Model 3	Model 4
Fixed Effect	Model estimates (standard error)				
Intercept (Y_{00})	2.953* (0.055)	2.934* (0.058)	2.819* (0.060)	2.930* (0.054)	2.819* (0.060)
DateCentred (Y_{10})		0.0003 (0.0003)	0.007* (0.001)	0.000* (0.000)	0.007* (0.001)
DateC_kwdr (Y_{01})			−0.000052* (0.000008)		−0.000* (0.000)
Random Part	Variance component (standard error)				
Intercept (τ_{00})	0.076* (0.022)	0.076* (0.022)	0.076* (0.022)	0.064* (0.020)	0.076* (0.022)
Residual (σ^2)	0.026* (0.003)	0.026* (0.003)	0.020* (0.002)	0.025* (0.003)	0.020* (0.002)
Slope (τ_{11})				0.000[a] (0.000)	0.000[a] (0.000)
Covariance (τ_{01})				1.000[a] (0.000)	−1.000[a] (0.000)
Model summary					
Deviance statistic	−66.409	−53.119	−68.991	−56.003	−68.991
# of parameters	3	4	5	6	7

*$p < 0.01$
[a]This covariance parameter is redundant. The statistic and confidence interval cannot be computed.

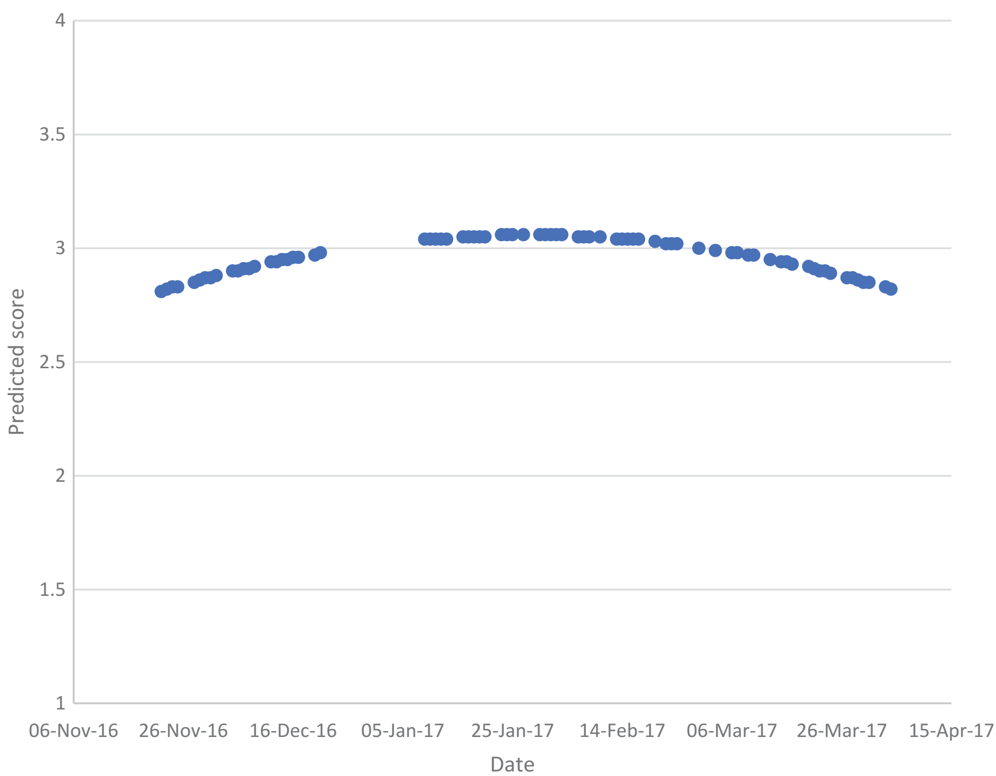


Figure 9. Predicted values of teachers' teaching quality plotted against date.

Conclusion and discussion

When using smartphone-assisted student feedback, teachers recognised where there is room for improving their lessons (hypothesis one confirmed). Teachers did not seem to reflect more on their lessons in response to the smartphone-assisted student feedback (hypothesis two rejected). As regards to hypothesis three (which can be confirmed), teachers in this study undertook improvement-oriented actions *outside* the lessons. They mostly reported that they have looked at the feedback, followed by 'I have discussed the feedback with students' and 'I have discussed the feedback with colleagues.' Teachers also undertook actions *during* the lessons, in response to the tool feedback. The number of actions varied per teaching aspect (e.g. actions related to the teaching 'clear instruction' were mentioned more than actions related to 'good student–teacher interaction'). Hypothesis four can be rejected, as teachers' teaching quality initially improved slightly, but after some time it capriciously dropped to the initial level.

Discussion

Although teachers as a result of the feedback obtained more insight into where they could improve their lessons, and although they reported improvement-oriented efforts in response to the student feedback, teachers did not improve their teaching quality sustainably. How could that happen? Several explanations for this finding can be considered.

First of all, as reflection is an important step on the road towards improvement (Driessen et al., 2008; Fraser, 2007; Hoban, 2004), it may be that, because teachers did not seem to reflect significantly more on their lessons, the actions conducted were not based on thorough problem analyses and therefore were not effective.

Also, it could be that the teaching aspects that were in need of improvement required the improvement of complex teaching skills (e.g. aspects of clear instruction, differentiation, and asking questions that cognitively activate students, Van de Grift, 2010) which cannot be improved by teachers on their own (Roche & Marsh, 2000). In this study, teachers were provided with feedback without any further improvement support as we wanted to study the effects of such an intervention.

Although there are factors influencing the use and effects of student feedback (mentioned in the theoretical framework of this study), it went beyond the scope of this publication to take them into account. They will be studied in the analyses for another article. However, the influencing factors may have a strong impact on the effects of smartphone-assisted student feedback on teachers. For example, if the school context is not supportive in using the feedback, or if teachers are not willing to change their way of teaching, then the feedback probably will remain unused and be ineffective.

Moreover, the research on deliberate practice by Ericsson (2006) shows that becoming an expert in a domain (someone consistently outperforming colleagues in a specific domain) requires a strong improvement motivation, stepping out of your comfort zone (searching for your weaknesses), clear definitions of ideal (expert) behaviour, the definition of small and precise improvement goals, and intensive practice until the goal set has been accomplished. It also requires that a coach knows what ideal behaviour (e.g. quality classroom management) looks like, how it can be trained effectively and which practices are effective if problems occur during the improvement process. It is impossible for a teacher to know and do all this on his/her own. Thus, the use of the Impact! tool can serve as a 'quick scan' for teachers that provides them with insight into where there is room for improvement. That is, however, only the starting point. An interesting question is how the other prerequisites for expertise development can be fulfilled in schools: how can the Impact! feedback to teachers be combined with the characteristics of deliberate practice in such a way that it matches with what is possible within the context of schools?

Limitations of the study and recommendations for further research

First, some practical conditions influenced the intervention in a negative way (Seifert, 2018), for example Internet problems, the need for technology support for (older) teachers, students using different types of mobile phones (different operating systems might make the tool work or not) and students being distracted (because of chatting and text messaging) when they were supposed to fill out the questionnaire.

A second remark on the limitations of the study relates to teachers' self-report for measuring several variables in this study (e.g. teachers' reflection on their lessons, their improvement-oriented actions). As self-reports may be susceptible to bias such as social desirability (Moorman & Podsakoff, 1992), they have their shortcomings (even if it was considered the best possible way of measuring).

Thirdly, teachers in the experimental group might have adapted some of their approaches because they had to fill out two extra questionnaires. However, considering the findings (no increased reflection on teaching quality, no sustained improvement of teaching quality), this effect may only have occurred for the improvement-oriented actions undertaken by teachers.

Another methodological remark refers to the fact that teaching quality (and the possible changes in teaching quality) was (were) measured by means of student perceptions. It may be that students have a fixed, subjective image of the quality of the lessons of their teachers that is hard to change. It could be that teachers actually did improve their teaching quality, but students did not see and report it (it could be, but we do not know). Further research is therefore required into the reliability and validity of student perceptions of (changes in) teaching quality, and on the factors influencing student perceptions (e.g. student, class and teacher characteristics). Student ratings of lessons in the case of an intervention might be compared with the scores of external observers of the same lessons, to study whether both actors notice the changes in teaching quality.

Disclosure statement

No potential conflict of interests was reported by the authors.

Notes on contributors

Hannah J.E. Bijlsma is a PhD candidate at the Department of Teacher Development ELAN of the University of Twente in the Netherlands and a primary school teacher (Grade 1). She obtained two Master degrees from the University of Groningen and is interested in teacher effectiveness and the use of student feedback for the improvement of teaching.

Adrie J. Visscher is a full professor at the University of Twente and holds an endowed chair in data-based decision making at the University of Groningen. He is interested in how teachers can be supported in optimising lesson quality and their impact on student learning by providing them with various types of feedback, and by training teachers for differentiating teaching in line with students' varying instructional needs.

Marjoleine J. Dobbelaer is a PhD candidate at the University of Twente. Her current research focuses on teacher evaluation by means of classroom observation. Her research interests lie in the area of teacher development, school improvement and accountability. Marjoleine graduated in educational science and worked as a researcher at the Dutch Inspectorate of Education with a focus on the quality of teachers and school leaders.

Bernard P. Veldkamp is head of the Department of Research Methodology, Measurement and Data Analysis at University of Twente and the scientific director of the Research Center for Examination and Certification. His research interests include measurement optimisation, behavioural data science and computerised assessment. His current projects are about optimising the use of various types of student information and about learning processes in relation to formative assessments.

ORCID

Hannah J.E. Bijlsma http://orcid.org/0000-0001-8825-0167
Adrie J. Visscher http://orcid.org/0000-0001-8443-9878

References

Balch, R. T. (2012). *The validation of a student survey on teacher practice*. Nashville, TN: Vanderbilt University.

Bell, L. M., & Aldridge, J. M. (2014). Investigating the use of student perception data for teacher reflection and classroom improvement. *Learning Environment Research*, *17*, 371–388.

Bijlsma, H. J. E. (2016). *Gathering feedback by means of student perceptions. A systematic review of student perception questionnaires* (Unpublished Master's thesis). University of Groningen, Groningen.
Black, P., & Wiliam, W. (2006). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practices*, *5*, 7–74.
Bron, R., Endedijk, M. D., Van Veelen, R., & Veldkamp, B. P. (2018). The joint influence of intra- and inter-team learning processes on team performance: A constructive or destructive combination. *Vocations and Learning*, 1–26. https://doi.org/10.1007/s12186-018-9197-z
Cantrell, S., & Kane, T. J. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Bill & Melinda Gates Foundation.
Cook-Sather, A. (2002). Authorizing students' perspectives: Toward trust, dialoque, and change in education. *Educational Researcher*, *31*(4), 3–14. doi:10.3102/0013189X031004003
Cook-Sather, A. (2007). What would happen if we treated students as those with opinions that matter? The benefits to principals and teachers of supporting youth engagement in school. *NASSP Bulletin*, *91*(4), 343–362. doi:10.1177/0192636507309872
COTAN. (2010). *Beoordelingssysteem voor de kwaliteit van tests. Geheel herziene versie [Quality of tests evaluation system. Completely revised version]*. Utrecht: NIP.
Creemers, B. M. P. (1994). *The effective classroom*. London: Cassell.
Day, C., Sammons, P., & Kington, A. (2008). *Effective classroom practice (ECP): A mixed-method study of influences and outcomes*. Nottingham: School of Education, University of Nottingham.
De Vocht, A. (2016). *Basisboek SPSS 24 [Handbook SPSS 24]*. Utrecht: Bijleveld Uitgeverij.
Den Brok, P., Brekelmans, M., & Wubbels, T. (2007). Multilevel issues in research using students' perceptions of learning environments: The case of the Questionnaire on Teacher Interaction. *Learning Environment Research*, *9*, 199–213.
Driessen, E., Van Tartwijk, J., & Dornan, T. (2008). The self-critical doctor: Helping students become more reflective. *BMJ*, *336*, 827–830.
Drijvers, P. (2015). *Denken over wiskunde, onderwijs en ICT*. Rede uit gesproken op 21 mei 2015 aan: de Universiteit Utrecht [Thinking about mathematics, education and ICT. Inauguration speech, 21 May 2015 at the University of Utrecht].
Dunn, T. J., Baguley, T., & Brunsden, V. (2013). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, *105*, 399–412. doi:10.1111/bjop.12046
Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In K. A. Ericsson (Ed.), *The Cambridge handbook of expertise and expert performance* (Vol. 38, pp. 685–705). New York, NY: Cambridge University Press.
Ertmer, P. A., & Newby, T. J. (1996). The expert learner: Stragic, self-regulated and reflective. *Instructional Sciences*, *24*, 1–24.
Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Students ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, *29*, 1–9.
Ferguson, R. F. (2012). Can student surveys measure teaching quality? *Phi Delta Kappan*, *94*(3), 24–28.
Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll*. Los Angeles, CA: Sage.
Fogg, B. J. (2009). *A behavioural model for persuasive design*. Stanford University: Persuasive Technology Lab.
Fraenkel, J. R., Hyun, H. H., & Wallen, N. E. (2012). *How to design and evaluate research in education*. New York, NY: Mcgraw-Hill Education.
Fraser, B. J. (1998). Science learning environments: assessment, effects and determinants. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 527–564). London: Kluwer Academic Publishers.
Fraser, B. J. (2007). Classroom learning environments. In S. K. Abell & N. G. Lederman (Eds.), *Handbook of research in science education* (pp. 103–124). Mahwah, NJ: Lawrence Erlbaum.
Fresko, B., & Nasser, F. (2001). Interpreting student ratings: consultation, instructional modification, and attitudes towards course evaluation. *Studies in Educational Evaluation*, *27*, 291–305.
Gaertner, H. (2014). Effects of student feedback as a method of self-evaluating the quality of teaching. *Studies in Educational Evaluation*, *42*, 91–99.
Hattie, J. (2008). *Visible learning. A synthesis over 800 meta-analyses relating to achievement*. Abingdon: Routledge.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487
Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371–404). Charlotte, NC: Information Age.
Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, *41*(2), 56–64.
Hoban, G. F. (2004). Enhancing action learning with student feedback. *Action Learning: Research and Practice*, *1*, 203–218.

Hoban, G. F., & Hastings, G. (2006). Developing different forms of student feedback to promote teacher reflection: A 10-year collaboration. *Teaching and Teacher Education*, *22*, 1006–1019.
Hollingsworth, J., & Ybarra, S. (2015). *Explicit direct instruction (ED): The power of the well-crafted, well-taught lesson*. Thousand Oaks, CA: Corwin Press.
Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, *64*, 349–371. doi:10.1037/0021-9010.64.4.349
Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper. MET project*. Seattle, WA: Bill & Melinda Gates Foundation.
Kember, D., Leung, D. Y. P., Jones, A., Yuen Loke, A., McKay, J., Sinclair, K., … Yeung, E. (2010). Development of a questionnaire to measure the level of reflective Thinking. *Assessment & Evaluation in Higher Education*, *25*(4), 380–395.
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254–284.
Könings, K. D., Van Berlo, J., Koopmans, R., Hoogland, H., Spanjers, I. A. E., Ten Haaf, J. A., & Van Merriënboer, J. J. G. (2016). Using a smartphone app and coaching group sessions to promote residents' reflection in the workplace. *Academic Medicine*, *91*, 365–370.
Lasagabaster, D., & Sierrra, J. M. (2011). Classroom observation: Desirable conditions established by teachers. *European Journal of Teacher Education*, *34*, 449–463.
Leatherdale, S. T., Cameron, R., Brown, K. S., & McDonald, P. W. (2005). Senior student smoking at school, student characteristics, and smoking onset among junior students: A multilevel analysis. *Prev. Med.*, *40*, 853–859.
Levin, J. A., & Datnow, A. (2012). The principal role in data-driven decision making: Using case-study data to develop multi-mediator models of educational reform. *School Effectiveness and School Improvement*, *23*, 179–202.
Lincoln, Y. S. (1995). In search of student voices. *Theory into Practice*, *34*, 88–93.
MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teachers. *Innovative Higher Education*, *40*, 291–303.
Marzano, R. J. (2003). *What works in schools. Translating research into action*. Alexandria, VA: ASCD.
Maulana, R., Helms-Lorenz, M., & Van de Grift, W. J. C. M. (2015). Development and evaluation of a questionnaire measuring pre-service teachers' teaching behaviour: A Rasch modelling approach. *School Effectiveness and School Improvement*, *26*, 169–194. doi:10.1080/09243453.2014.939198
Miller, M. T. (1971). Instructors' attitudes toward, and their use of, student ratings of teachers. *Journal of Educational Psychology*, *62*, 235–239.
Moorman, R. H., & Podsakoff, P. M. (1992). A meta-analytic review and empirical test of the potential confounding effects of social desirability response sets in organizational behaviour research. *Journal of Occupational and Organizational Psychology*, *65*, 131–149.
Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, *12*, 53–75.
Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B. M. P., Timperley, H., & Earl, L. (2014). State of the art – Teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, *25*, 231–256.
Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, *24*, 485–500.
Pambookian, H. S. (1972). *The effect of feedback from students to college instructors on their teaching behavior* (Unpublished doctoral dissertation). University of Michigan.
Pambookian, H. S. (1976). Discrepancy between instructor and student evaluations of instruction: Effect on instructor. *Instructional Sciences*, *5*, 63–75.
Peterson, K. D., Wahlquist, C., & Bone, K. (2000). Student surveys for school teacher evaluation. *Journal of Personnel Evaluation in Education*, *14*, 135–153.
Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, *48*, 85–112. doi:10.1016/j.jsp.2009.09.002
Pianta, R. C., & Hamre, B. K. (2006). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*, 109–119.
Quaglia, R., & Corso, M. (2017). *Lef om te luisteren. Met leerlingen werken aan de kwaliteit van onderwijs [Student voice: The instrument of change]*. Huizen: Pica.
Reynolds, D., Sammons, P., De Fraine, B., Van Damme, J., Towsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement*, *25*, 197–230.
Roche, L. A., & Marsh, H. W. (2000). Multiple dimensions of university teacher self-concept: Construct validation and the influence of students' evaluations of teaching. *Instructional Sciences*, *28*, 439–468.
Rosenshine, B. (1995). Advances in research on instruction. *Journal of Educational Research*, *88*, 262–268.
Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research*. London: Institute of Education.
Santos, J. R. A. (1999). Cronbach's alpha: A tool for assessing the reliability of scales. *Tools of the Trade (Texas A&M University), 37*(2). Retrieved from https://joe.org/joe/1999april/tt3.php

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *Measuring Reform Practices*, *102*, 245–253.

Schildkamp, K., & Lai, M. K. (2013). Conclusions and a data use framework. In K. Schildkamp & M. K. Lai (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 177–191). Dordrecht: Springer.

Seifert, T. (2018, March). *Smartphone implementation from lecturers, teachers, students and pupils' perspectives*. Paper presented at the SITE conference, Washington, DC.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*, 153–189.

Sijtsma, K. (2009). Over misverstanden rond Cronbach's alpha en de wenselijkheid van alternatieven [About misunderstandings of Cronbach's alpha and the desirability of alternatives]. *De Psycholoog [The Psychologist]*, *44*, 561–567.

Sluijsmans, D., & Kneyber, R. (2016). *Toetsrevolutie. Naar een feedbackcultuur in het voortgezet onderwijs [Tests revolution. Towards a feedback culture in secondary education]*. Leiden: Phronese.

Snijders, T., & Bosker, B. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Timmers, C. F., & Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers in Education*, *56*, 923–930.

Tuckman, B. W., & Olivier, W. F. (1968). Effectiveness of feedback to teachers as a function of source. *Journal of Educational Psychology*, *59*, 297–301.

Van de Grift, W. J. C. M. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. *Educational Research and Evaluation*, *49*, 127–152.

Van de Grift, W. J. C. M. (2010). *Ontwikkeling in de beroepsvaardigheden van leraren*. Rede uitgesproken op 23 maart 2010: Rijksuniversiteit Groningen [Development in the professional skills of teachers. Inauguration speech, 23 March 2010: University of Groningen].

Van der Lans, R. M. (2017). *Teacher evaluation through observation* (Unpublished doctoral dissertation). University of Groningen.

Van der Lans, R. M., Van de Grift, W. J. C. M., & Van Veen, K. (2016). Developing a teacher evaluation instrument to provide formative feedback using students' ratings of teaching acts. *Educational Measurement: Issue and Practice*, *34*(3), 18–27.

Van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. (in press). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement*.

Van der Scheer, E. A., & Visscher, A. J. (2017). Effects of a data-based decision-making intervention for teachers on students' mathematical achievement. *Journal of Teacher Education*, 1–14. doi:10.1177/0022487117704170

Visscher, A. J. (2015). *On the relevance of data-based decision-making in education*. Inaugural speech, University of Groningen.

Voerman, L., Meijer, P. C., Korthagen, F. A. J., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, *28*, 1107–1115.

Weisberg, H. F. (1992). *Central tendency and variability*. Sage University paper series on Quantitative Applications in the Social Sciences. Newbury Park, CA: Sage.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: Applying the principal-agent framework. *School Effectiveness and School Improvement*, *19*, 239–259.