Do teachers know how their teaching is perceived by their pupils?

Pham, Giang; Koch, Tobias; Helmke, Andreas; Schrader, Friedrich-Wilhelm; Helmke, Tuyet; Eid, Michael

Published in:
Procedia - Social and Behavioral Sciences

DOI:
10.1016/j.sbspro.2012.06.068

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 46 (2012) 3368 – 3374

Procedia
Social and Behavioral Sciences

WCES 2012

Do teachers know how their teaching is perceived by their pupils?

Giang Pham [a] *, Tobias Koch [b], Andreas Helmke [c], Friedrich-Wilhelm Schrader [c], Tuyet Helmke [c], Michael Eid [b]

[a] *Graduate School of Learning and Teaching Processes "Upgrage", University of Koblenz-Landau, Fortstraße 7, 76829 Landau, Germany*
[b] *Department of Educational Science and Psychology, Freie Universität Berlin, Habelschwerdter Allee 4, 14195 Berlin, Germany*
[c] *Department of Psychology, University of Koblenz-Landau, Fortstraße 7, 76829 Landau, Germany*

Abstract

Research findings have revealed that teachers' perceptions of their own instructional quality are often inappropriate. Our research dealt with two research questions. First, how closely are teachers' ratings of instruction associated with their students' ratings? Second, do teachers share a stronger consensus with their students than visiting colleagues do? Data were collected by means of questionnaires and analyzed using the correlated-trait - correlated-method minus one (CTC(M-1)) model (Eid, 2000). Our results show low consistency between teacher and student ratings, which is not higher than between students and colleagues. Consequences for research and practice are discussed.

Keywords: classroom instruction, perception of instruction, consistency between perspectives, CTC(M-1);

1. Introduction

How school education can be improved effectively is an ongoing issue for both researchers and practitioners. According to the currently most comprehensive meta-analysis of over 800 meta-analyses relating to students' learning achievement, teachers' proficiency and instructional quality are among the most powerful influences (Hattie 2009, p.238). It is concluded that it would be an "excellent beginning" to make teaching progress if teachers - as activators of students' learning - were able to see learning "through the eyes of students" (Hattie 2009, p.252), since students themselves, not teachers, decide what and how they learn (Olson 2003). Nevertheless, research findings on classroom instruction have revealed that teachers' perceptions are often quite different from the view of their pupils, whereas academic achievement has been much better predicted by students' view of instructional quality (Fraser 1991; Clausen 2002; Kunter and Baumert 2006). Furthermore, findings from video studies have shown a big gap between teachers' self-judgment of own speaking time and their real speaking time, which was exactly measured using the recorded videos (Helmke, Helmke et al. 2008). Considering the significance of the problem, we have developed a nationwide program named EMU (evidence-based methods of diagnosis of classroom instruction), which teachers may use to compare their own view of instruction with external views (Helmke, Helmke et al. 2011). The program instruments consist of a basic brochure, additional text (via hyperlink), questionnaires and software. The brochure contains helpful tips and links for teachers regarding why and how they should conduct the survey, how the results can be interpreted and what should be implemented to improve their instructional quality. The software automatically generates graphic results after data entry. Questionnaires are used

* Giang Pham. Tel.: +49-6341-28031-222
E-mail address: giang@uni-landau.de

1877-0428 doi:10.1016/j.sbspro.2012.06.068

to obtain the data. EMU includes ratings from two external perspectives: students and visiting teachers (colleagues). In those aspects of classroom instruction, where students don't have the necessary resources (e.g., didactical knowledge) to evaluate comprehensively, teachers may benefit more from a comparison between self-judgment and that of a visiting colleague ("critical friend"). With EMU, we aim at helping teachers to acquire a differentiated, data-based feedback of their own classroom instruction in order to improve teaching quality by means of cooperative efforts. Furthermore, EMU data help us to obtain a deeper view into the visibility of teaching and learning. The first question we investigate is whether teachers can rate their instructional quality appropriately as it appears in the eyes of their students. Since the ratings of visiting colleagues are available, we also conduct an analysis to find out if teachers share a stronger consensus with their students than their colleagues do.

2. Method

2.1. Research instruments

Data were collected by means of questionnaires with equivalent items for teachers, students and visiting colleagues to enable a comparison. As colleagues can only see one lesson they visit, all raters are requested to judge just one concrete lesson. Analogously, as students can only perceive own feelings, own thoughts, student items are formulated from the "I"-perspective. All item formulations are student-centered, so that teacher and colleague ratings are about how the lesson was perceived by students. Since the instruments primarily serve as self-reflection tools for teachers and are applied in the classroom context, they should be economic and the rating-indicators must be observable for all users. Furthermore, the instruments should be applicable in different school types and subjects. Hence, among a wide range of dimensions of instructional quality and their indicators, our main research area contains a scale for classroom outcomes as well as four quality dimensions, which are empirically confirmed as significant and influential for students' outcomes in Germany. These are classroom management (CM, 5 items), learning climate (CL, 7 items), clarity and structuring (ST, 6 items) and activation (7 items) (e.g., Helmke 2010, Borich 2007). Nevertheless, the instruments are not limited to these four dimensions and the outcome scale, as teachers can optionally add more items to the survey in their classroom by filling in a so-called wild-card area in the questionnaires. All items have a 4-point response format (disagree to agree). For research analysis we use items from the first three dimensions, because different lesson types (e.g., plenum, group work…) as well as different school subjects (e.g., sport, mathematics…) are available, which can to a large extent influence the formal activities of lessons.

2.2. Participants and procedure

The study comprised participants from 112 classrooms (5^{th} to 13^{th} grade) from different types of schools in 5 states in Germany (Baden-Württemberg, Hessen, Niedersachsen, Nordrhein-Westfalen, Sachsen-Anhalt). In total there were 112 teacher tandems (1 teacher and 1 visiting colleague each) and 2513 students from 111 classrooms (in one classroom in primary school there were only teacher's and colleague's data). It was not a representative sample. Teachers conducted the survey themselves, following the guidance in the EMU brochure. After data entry in the software, teachers retrieved the results for their own analyses; they also exported their data and sent us the generated anonymous data files. Data from each classroom were identified with a teacher's ID, which was personally defined by the teacher and consisted of ten digits.

2.3. Analysis model

2.3.1. The correlated-trait – correlated method minus one (CTC(M-1)) model

The measurement design plays an important role in modeling multitrait-multimethod (MTMM) data. Eid, Nussbeck, Geiser, Cole, Gollwitzer, & Lischetzke (2008) distinguish three measurement designs: (1) measurement designs using structurally different methods, (2) measurement designs using interchangeable methods, and (3)

measurement designs incorporating both types of methods. It is important to note that each measurement design implies a different data structure, and therefore a different MTMM model (see Eid et al., 2008). Interchangeable methods can be conceived as methods that are randomly drawn out of a set of equivalent methods (e.g., student ratings). In contrast, structurally different methods are methods which are fixed beforehand for each target (e.g., teacher self-ratings and corresponding colleague ratings). Whereas measurement designs for interchangeable methods imply a multilevel data structure (raters nested within targets), structurally different methods do not imply this kind of clustering. Eid (2000, see also Eid et al. 2003) proposed a correlated-trait - correlated-method minus one (CTC(M-1)) model, which enables researchers to contrast different methods against a reference (standard) method. Within the CTC(M-1) framework method, effects are defined as residuals. Thus, researchers are able to investigate the over- or underestimation of the reference (standard) method (e.g., self-ratings) with a non-reference method (e.g., teacher ratings). Eid et al. (2008) presented an extended CTC(M-1) model, which can be used for measurement designs with interchangeable and structurally different methods. In this study we used the extended (multilevel) CTC(M-1) model for analyzing the data. Due to our research questions as well as the higher reliability of students' ratings than those of teachers and colleagues, we selected the student ratings as reference (standard) method. Teacher and colleague ratings were chosen as non-reference method. The total measurement equation for an indicator pertaining to a reference method (student rating) is given by:

$$Y_{rtij1} = \alpha_{\mathrm{T}ij1} + \lambda_{\mathrm{T}ij1} * T_{tij1} + \lambda_{\mathrm{UM}ij1} * UM_{rtj1} + \varepsilon_{rtij1}$$

The subscript *r* denotes a particular student (interchangeable rater) who rates a target *t* (teacher/class) on an indicator *i* measuring a construct *j* with a method *k* (in this case the reference method, *k*=1). Note that, all factor loadings pertaining to a particular method factor are fixed to 1. With respect to this measurement equation, it is clear that there is only a method factor pertaining to the reference method on level-1. The unique method factor UM_{rtj1} reflects the over- or underestimation of the true group mean of student ratings in a class. T_{tij1} represents the true group mean of student ratings (trait factors) in a class, which is weighted by $\lambda_{\mathrm{T}ij1}$ factor loading; $\alpha_{\mathrm{T}ij1}$ is a constant intercept, and ε_{rtij1} is the measurement error. The total measurement equation for an indicator pertaining to the non-reference method (teacher and colleague rating) is given by:

$$Y_{tijk} = \alpha_{\mathrm{T}ijk} + \lambda_{\mathrm{T}ijk} * T_{tij1} + \lambda_{\mathrm{M}ijk} * M_{tijk} + \varepsilon_{tij1}, \forall k \neq 1$$

Note that the true group means of the student ratings are taken as predictors for the true scores of the remaining methods (teacher and colleague ratings). Therefore, the method factors on level-2 are also defined as residuals with respect to the latent regression described above. Consequently, all method factors have an expected value of zero and are uncorrelated with the reference method factor. Due to the additive decomposition of the latent variables in the measurement equations, it is possible to define coefficients of consistency and method specificity.

The consistency coefficient for the true variance of indicator Y_{tijk} of the non-reference methods (teacher and colleague ratings) is given by following equation:

$$Con\left(Y_{tijk}\right) = \frac{\lambda^2_{\mathrm{T}ijk} * Var(T_{tij1})}{\lambda^2_{\mathrm{T}ijk} * Var(T_{tij1}) + \lambda^2_{\mathrm{M}ijk} * Var(M_{tjk})}$$

This coefficient represents the amount of true variance of a non-reference variable that is explained by the reference trait factor (T_{tij1}). The square root of this coefficient can be interpreted as the convergent validity between a non-reference method (true teacher or colleague rating) and the reference method (true average student rating).

The method specificity coefficient of the true variance of the non-reference variables is defined by:

$$Spe(Y_{tijk}) = \frac{\lambda^2_{\mathrm{M}ijk} * Var(M_{tjk})}{\lambda^2_{\mathrm{T}ijk} * Var(T_{tij1}) + \lambda^2_{\mathrm{M}ijk} * Var(M_{tjk})}$$

The method specificity coefficient of the non-reference methods represents the amount of true variance that is not explained by the reference trait factor. Hence, this coefficient is an indicator for method specificity. With regard to our main research questions, we are interested in the extent to which the teacher ratings can or cannot be predicted by the reference method (true average student ratings). In addition to that, we are interested in whether teachers have more in common with their students than their colleagues do. In other words, to what extent do the consistency coefficients differ between indicators pertaining to teacher or colleague ratings?

2.3.2. Model specification

We used two item parcels as indicators for each trait-method-unit (TMU). There were two reasons for using item parcels instead of the raw items. First, the items contain only 4 categories and are slightly skewed. Thus, a model for categorical items would have been required. However, the estimation process (using numerical integration) for this kind of multilevel structural equation model with categorical items would have been too demanding. Second, there are too many items (5-7) per TMU to identify all parameters. This model could only be identified with two item parcels per TMU. For identification purposes we restricted the factor loadings for each method factor to be equal. In addition, we specified item-specific reference traits, given that both item parcels do not necessarily need to be homogenous.

This model is depicted in Figure 1. The analysis was conducted using Mplus 5.0 and maximum likelihood estimator with robust standard errors (MLR) (Muthén and Muthén 2007).

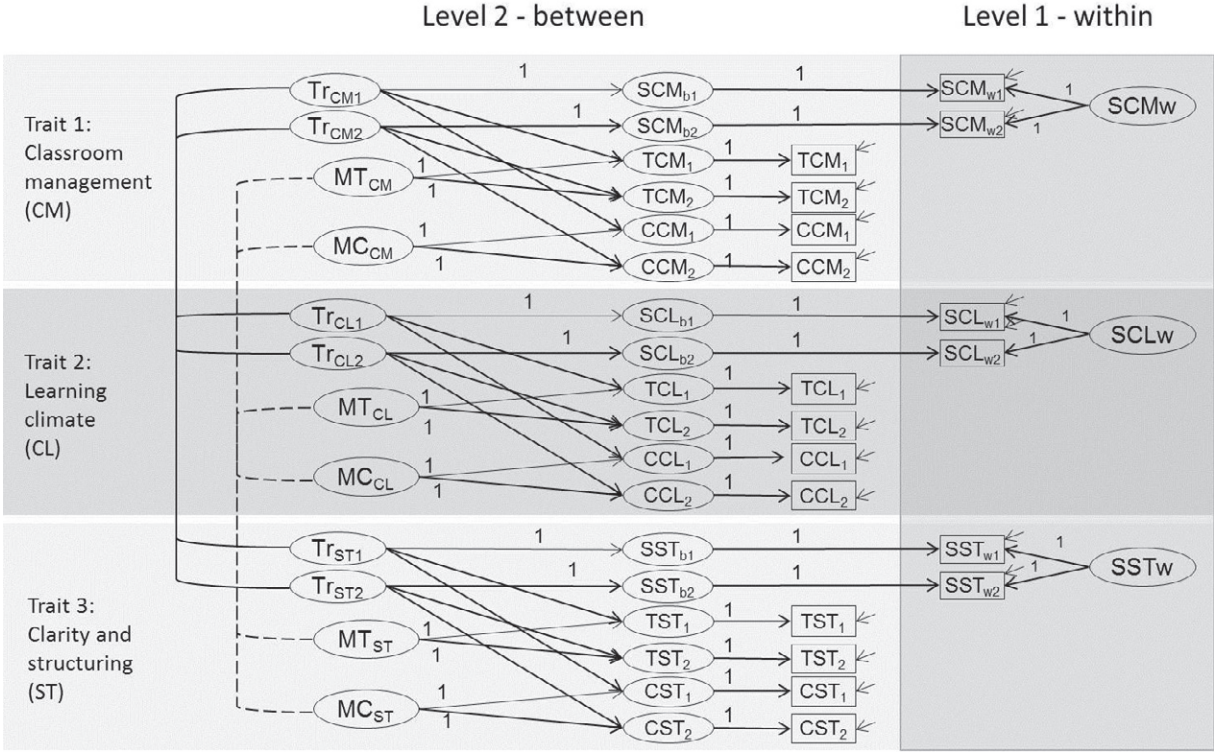

Figure 1: Multilevel CFA-CTC(M-1) model of instructional quality for structurally different and interchangeable methods.

Note. Tr=Trait, MT=method factor for teacher ratings, MC=method factor for colleague ratings, S=student ratings, b=between, w=within, T=teacher ratings, C=colleague ratings, subscript (*i*)=indicator, *i*=1=first item parcel, *i*=2=second item parcel. All fixed loading parameters are indicated with real numbers. The curved lines (full and dashed) between the constructs indicate that all of the linked constructs can be correlated.

3. Results

The illustrated model in Figure 1 fits the data well: χ^2(df=114, N=2453) = 193.259, $p < 0.01$, CFI=.98, TLI=.97, RMSEA=.02, SRMR for within=.01, SRMR for between=.08.

Table 1: Mean ratings of students, teachers and colleagues

Indicator	Means S	Means T	Means C
CM1	3.18	3.05	3.27
CM2	3.34	3.21	3.43
CL1	3.15	3.10	3.31
CL2	3.25	3.06	3.23
ST1	2.96	2.87	3.10
ST2	3.16	2.73	3.06

Note: Means=means of indicator, S=student ratings, T=teacher ratings, C=colleague ratings, Min=1, Max=4

Table 1 shows the estimated means of indicators rated by students, teachers and colleagues sequentially. On average, students and colleagues rated the instructional quality higher than teachers did. That means teachers largely underestimated their classroom instruction in comparison to other raters. It could be hypothesized that the underestimation was due to a generally high (self-) expectation of teachers. However, we still did not know if the variation between teacher ratings was in line with the variation between the ratings of students in their classrooms. Likewise, it had not yet been shown if teacher and student data matched better than that of students and colleagues.

Table 2: Reliability, method specificity, and consistency coefficients estimated for teacher ratings

Indicator	Rel	Spe	Con
TCM1	0.69	0.59	0.41
TCM2	0.45	0.84	0.16
TCL1	0.66	0.64	0.36
TCL2	0.63	0.67	0.35
TST1	0.60	0.85	0.15
TST2	0.62	0.71	0.29

Note: T=teacher ratings, Rel=reliability, Spe=method specificity, Con=consistency

Table 2 presents the reliability, consistency and specificity coefficients of teacher variables. As shown in this table, the reliability (Rel) of teacher indicators was acceptable (Rel>=0.60) except for the second indicator of CM (Rel=0.45). The low reliability coefficients of the items might be explained by the heterogeneity of the items. With regard to the extent to which teacher ratings were explained by model factors, it can be seen in Table 2 that the true expected student ratings did not predict or correlate highly with those of teachers. Concretely, the consistency coefficients - the percentage of explained variance of the true expected teacher ratings by trait factors (Tr-variables in model in

Figure 1) - ranged from 15% to 41%. The correlations between the teacher and the student ratings are the roots of the consistency coefficients, which ranged from 0.39 to 0.64. These were somewhat higher than the ones reported in previous research, due to the correction of the measurement error (latent correlation). Nevertheless, the low consistency implies that a larger amount of variance of the teacher ratings was not shared with the student ratings, but was specific to teachers (method specificity). Teacher method specificity coefficients (Spe = 100% - consistency) ranged from 59% to 85%, which was much higher than the consistency for all indicators. Overall, the true scores of teacher ratings of classroom instruction cannot to a large extent be explained by the true student perceptions.

Table 3: Reliability, method specificity, and consistency coefficients estimated for colleague ratings

Indicator	Rel	Spe	Con
CCM1	0.41	0.51	0.49
CCM2	0.33	0.39	0.61
CCL1	0.57	0.54	0.46
CCL2	0.46	0.67	0.35
CST1	0.64	0.77	0.23
CST2	0.54	0.72	0.28

Note: C=colleague ratings, Rel=reliability, Spe=method specificity, Con=consistency

The second question we were interested in was whether the consistency coefficients of teacher ratings were higher than those of colleague ratings. Data shown in Table 3 do not confirm that this was the case. The consistency coefficients of colleague ratings ranged from 23% to 61% (latent correlations between student and colleague ratings were 0.48 to 0.78 respectively); they were even higher than the results of teacher ratings (15% to 41%). However, taking into account that the reliabilities of colleague ratings were low (0.33 to 0.64), a statement that colleagues show more consensus with students in ratings is not very convincing.

4. Discussion

In summary, teachers underestimated their quality of teaching in comparison to student and colleague ratings on average. Moreover, teachers showed a low consistency with students with regard to their ratings. In addition, their ratings did not generally correspond better than colleague ratings with the students' judgment of instructional quality. To conclude, teachers obviously do not know well how their teaching is perceived by their pupils. Thus, it is recommended that teachers should be made more aware of this problem, by providing opportunities to learn from external feedback (by students as well as by colleagues) and by helping them to explore different views of instruction and constructively reflect on these differences. This may serve as a starting point for improving teaching.

In general, our results confirmed previous findings on differences between perspectives in rating classroom instruction, although we found higher correlations between perspectives because measurement errors were eliminated from estimation. Once again, the result pattern underlines the validity problem of research on instructional quality using questionnaires. Furthermore, the size of correlations levels between perspectives differs, depending on dimensions and empirical indicators. Thus, researchers should carefully consider which perspective should be chosen as the basis for research on classroom instructional quality.

References

Borich, G. D. (2007). Effective teaching methods. Research-based practice. Upper Saddle River, NJ, Pearson Education.

Clausen, M. (2002). Unterrichtsqualität: Eine Frage der Perspektive? Münster, Waxmann.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. Psychometrika 65(2): 241-261.

Eid, M., Lischetzke,T., Nussbeck, F. W. & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait-multimethod models: A multiple-indicator CTC(M-1) model. Psychological Methods 8(1): 38-60.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M. & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. Psychological Methods 13: 230-253.

Fraser, B. (1991). Two decades of classroom environment research. Educational environments: Evaluation, antecedents and consequences. H. J. Walberg. Elmsford, NY, Pergamon Press: 3-27.

Hattie, J. A. C. (2009). Visible Learning. A synthesis of over 800 meta-analyses relating to achievement. London, Routledge.

Helmke, A. (2010). Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts. Seelze, Klett-Kallmeyer.

Helmke, A., Helmke, T., Lenske, G., Pham, G., Praetorius, A.-K., Schrader, F.-W. & Ade-Thurow, M. (2011). EMU - Unterrichtsdiagnostik. Studienbrief Version 3.2. Kultusministerkonferenz: Projekt EMU (Evidenzbasierte Methoden der Unterrichtsdiagnostik). Landau, Universität Koblenz-Landau, Campus Landau.

Helmke, T., Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Die Videostudie des Englischunterrichts. Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie. DESI-Konsortium. Weinheim, Beltz: 345-363.

Kunter, M. & Baumert, J. (2006). "Who is the Expert? Construct and Criteria Validity of Student and Teacher Ratings of Instruction." Learning Environments Research 9(3): 231-251.

Muthén, L. K. & Muthén, B. O. (2007). Mplus User's Guide. Fifth Edition. Los Angeles, CA, Muthén & Muthén.

Olson, D. R. (2003). Psychological theory and educational reform: How school remakes mind and society. Cambridge, Cambridge University Press.