

The consistency between professed teaching practices and assessment practices: A case in mathematics class

Yoppy Wahyu Purnomo¹, Amrita Kaur, Siti Noor Binti Ismail², Didi Suryadi³, Sutawanir Darwis⁴

Abstrak: Penelitian ini bertujuan untuk (1) mengembangkan sebuah skala untuk mengukur praktik mengajar dan penilaian di kelas matematika; (2) mengidentifikasi profil praktik mengajar dan praktik penilaian di kelas matematika; dan (3) mengkaji konsistensi antar faktor-faktor yang bersesuaian pada praktik mengajar dan penilaian di kelas matematika. Metode survei *cross-sectional* dan studi pengembangan skala digunakan untuk menggapai tujuan tersebut. Partisipan dalam penelitian ini mencakup dua kelompok sampel guru sekolah dasar di Jakarta. Sampel pertama adalah 252 guru dan sampel kedua adalah 325 guru. Temuan penelitian ini mendapati terdapat dua faktor di setiap dimensi praktik guru di kelas matematika. Praktik mengajar mencakup faktor praktik relasional dan instrumental, sedangkan praktik penilaian mencakup faktor praktik *assessment for learning* dan *assessment of learning*. Penelitian ini juga menemukan bahwa sebagian besar guru mengarah pada praktik-praktik tradisional baik praktik mengajar maupun praktik penilaian, dan terdapat konsistensi antar faktor yang bersesuaian pada praktik mengajar dan penilaian tersebut. Diskusi lebih detail dapat ditemukan pada bagian hasil dan pembahasan.

Kata kunci: *Praktik penilaian, Praktik mengajar, Kelas matematika, Studi pengembangan skala, Survei cross-sectional*

Abstract: The purpose of this research was to (1) develop scales for measuring teaching practices and assessment practices in mathematics class; (2) identify the profile of teaching practices and assessment practices in mathematics class, and (3) examine the consistency between relevant factors in teaching practices and assessment practices in mathematics class. The methods of cross-sectional surveys and a scale development study were used to achieve that purpose. The participants in this research included two sample groups of primary school teachers in Jakarta. The first sample consisted of 252 teachers and the second sample consisted of 325 teachers. This research found that there were two factors in each dimension of teachers' practices in mathematics class. Teaching practices included the relational and instrumental practices, while the assessment practices included the assessment for learning and assessment of learning. This research also found that most of the teachers leaned toward traditional practices for both teaching practices and assessment practices and that there was a consistency between relevant factors in those teaching and assessment practices. A more detailed discussion can be found in the findings and discussion section.

Keywords: *Assessment practices, Teaching practices, Mathematics class, Scale development study, Cross-sectional Survey*

¹ Universitas Muhammadiyah Prof. DR. HAMKA, Jakarta, Indonesia, yoppy.w.purnomo@uhamka.ac.id

² Universiti Utara Malaysia, Malaysia

³ Universitas Pendidikan Indonesia, Bandung, Indonesia

⁴ Institut Teknologi Bandung, Indonesia

A. Introduction

Assessment has an inseparable role in the process of teaching and learning. This particular statement can often be found in scientific articles about the assessment of education (see Black & Wiliam, 1998; Kasih & Purnomo, 2016; Suci & Purnomo, 2016). The meaning of “inseparable” does not only mean that each learning requires assessment, but also that it has a role in continuing to fix and improve the process of learning.

Ever since Black and Wiliam's (1998) seminal paper, research regarding assessment practices integrated in mathematics learning became the focus of discussion in the last few decades (Balan, 2010; Kasih & Purnomo, 2016; Ma, Millman, & Wells, 2008; Mansyur, 2011; Purnomo, 2013, 2015, 2016b; Setiani, 2011). As an example, in the context of Indonesia, Setiani (2011) developed an alternative assessment in realistic mathematics learning, then Purnomo (2015) who was inspired by the works of Black and his colleagues (Black, Harrison, Lee, Marshall, & Wiliam, 2003, 2004; Black & Wiliam, 1998) to develop assessment-based learning in mathematics class. In his research, Purnomo (2015a) identified that an increase in students' mathematics performance could be optimized by emphasizing on the purpose and criteria of successful learning, the use of effective questioning strategy, constructive feedback, and giving the opportunity of self and peer assessment to students. The results of these research became valuable assets in providing theoretical and practical knowledge, to change the perception of teachers regarding assessment and in turn, could be implemented by them in mathematics class.

Utilization of research results is something that wants to be achieved in every research, in which the results of the research could be accepted, implemented and developed by the people utilizing those results (Cherney, Povey, Head, Boreham, & Ferguson, 2012; Lysenko, Abrami, Bernard, Dagenais, & Janosz, 2014). Therefore, it is of importance to examine how teachers' practice profiles are as a control toward the usefulness of and for research. However, empirical data related to teachers' practice profiles based on research in a mathematics class in Indonesia's context is rarely found in the literature. Furthermore, instruments to measure teachers' practices (e.g., assessment practices) in a mathematics class in Indonesia's context is also rarely found in the literature.

While assessment is an integrated part in the practice of teaching mathematics, it is also essential to see the consistency between assessment practices and teaching practices. It can be seen as stated by Purnomo (2016a) that for assessment practices emphasizing on the extent to which information is taken in by students after learning, the teachers' teaching will focus on how to lead students toward success in the criteria of standardized assessment. As for the other way around, when teaching practices focus more on student performance, then the assessment practices tend to lean toward the process of seeing the extent to which knowledge is taken in by students after learning. On the other hand, if teaching practices tend to give opportunities for students to construct their knowledge, then the assessment practices tend to be used for directing and guiding the process of learning and teaching through formative feedback. As for the other way around, when learning practices tend to be used for fixing the process of learning and teaching, then the teaching practices tend to be more relevant with the context and issues in class (Delandshere, 2002; Delandshere & Jones, 1999; Purnomo, Suryadi, & Darwis, 2016).

In their study, Wijaya, van den Heuvel-Panhuizen, and Doorman (2015) have examined the teaching practices of junior secondary mathematics teachers by use of observations, videos, and self-reports. However, this research only focused on teaching practices related to context-based mathematics learning. The use of those instruments was also limited to a small sample and more

interpretative results. Several researchers (Purnomo, 2017a; Swan, 2006) agreed that the use of questionnaires is critical to strengthen evaluation tools and also as a reflection of practices in mathematics class, especially for a larger sample.

The purpose of this research was to achieve answers to these questions: (1) How is the structure of factors underlying the teaching and assessment practice scales in a mathematics class? (2) What is the profile of teaching practices and assessment practices in mathematics class of the teachers in this research's sample? (3) Is there consistency between relevant factors in teaching practices and assessment practices?

B. Methods

This research used cross-sectional surveys and included a scale development study to achieve its purpose. The steps of the scale development study were adopted from Purnomo (2017a) which included (a) defining and specifying the measured construct, (b) developing the item pool, (c) providing and taking heed experts' corrections of the item pool (d) perfecting and validating the scale, and (e) evaluating the items.

The participants in this research included two sample groups. The first sample was for exploratory factor analysis purposes and the second sample was for confirmatory factor analysis purposes and also to answer the research's questions. The first group included 252 primary school teachers in Jakarta. They comprised of 200 females, 50 males, and two unidentified participants. Twenty-one percent of them had three years or less of teaching experience, 25,4% had 4-10 years of experience, 20% had 11-20 years of experience, and 32,5% had more than 20 years of experience. The second group included 325 primary school teachers in Jakarta, consisting 80,9% of females. Both samples represented several ethnic groups in Indonesia, such as the Javanese, Sundanese, Betawi, Minang, Melayu, Bima, Dayak, and several others. The participants from both samples were chosen through convenience sampling. Both sample sizes consisted of more than 200 participants, more than the recommended minimum sample size for factor analysis (Barrett, 2007; Comrey & Lee, 1992; Fabrigar, MacCallum, Wegener, & Strahan, 1999; Pituch & Stevens, 2016).

The instrument used in this research were questionnaires classified into three parts: questions about the teachers' demographic data, a questionnaire about teachers' teaching practices in mathematics class, and a questionnaire about teachers' assessment practices in mathematics class. The questionnaire was presented in Bahasa Indonesia. The scale's range used 5 points, which were always, often, sometimes, rarely, and never. The questionnaire about teachers' teaching practices in mathematics class was adapted from several items developed by Swan (2006), and the researchers themselves developed the rest. The questionnaire about teachers' assessment practices was adapted from several items developed by James et al. (2006), and the researchers themselves developed the rest.

After compiling the questionnaire's items, face and content validity were conducted qualitatively by two mathematics education experts, one research and evaluation of education expert, and two experienced teachers. They assessed the proposed questionnaire items presented in a booklet. This booklet included a description of the research's purpose, a summary of related literature, the definition of the measured construct, and instructions on how to fill in the sections regarding construct relevancy, clarity of a sentence, code of the problems, and the comment section. The draft of the instrument, after validation and suggestions from the expert team,

consisted of 16 items about teaching practices and 12 items about assessment practices. In total, the questionnaire consisted of 28 items.

Data analysis began with the use of principal factor analysis (PCA). After the PCA, analysis results of each factor were interpreted and linked to related literature. The analysis was continued with examining the internal consistency of the subscales through Cronbach's Alpha test with the help of SPSS statistical software version 21. The developed construct from results of the PCA was then validated by use of confirmatory factor analysis (CFA) with the help of the statistics software, SPSS Amos version 22. In literature, there are several researchers (e.g., Koh, Chai, & Tsai, 2010; Berg, 2008) that only reported their scale development studies using EFA or just CFA. This research used the criteria of Chi-Square statistic, the degree of freedom and p-value to be reported. In addition to that, normed Chi-Square (NC), root mean square error of approximation (RMSEA), normed-fit index (NFI), comparative fit index (CFI), and Tucker-Lewis index (TLI) were also reported using statistical criteria with a threshold in each index which are summarized in Table 1 (Purnomo, 2017a).

Table 1. Threshold measures for fit indices model

Index	Threshold	
	Good fit	Acceptable fit
NC	$1 \leq NC \leq 2$	$2 < NC \leq 3$
RMSEA	≤ 0.05	$0.05 < RMSEA \leq 0.08$
SRMR	≤ 0.05	$0.05 < SRMR \leq 0.08$
GFI	≥ 0.95	≥ 0.90
AGFI	≥ 0.95	≥ 0.90
NFI	≥ 0.95	≥ 0.90
CFI	≥ 0.95	≥ 0.90
TLI	≥ 0.95	≥ 0.90

Based on each formed construct, the analysis was continued with testing if the construct had adequate validity and reliability. Construct validity can be measured by different types of validity, such as face validity, content validity, concurrent validity and predictive validity, and convergent validity and discriminant validity (Drost, 2011). For this part, the analysis focused on convergent validity and discriminant validity. On the other hand, instrument reliability was shown by how well the coefficient of internal consistency was. Other alternatives can be found in the literature concerning how to examine convergent and discriminant validity. Convergent validity can be measured by how well the coefficients of the standardized factor loading, composite reliability (CR), and average variance extracted (AVE) are generated, while discriminant validity can be measured by comparing the square root of AVE for any two constructs and the correlation estimate between the same construct (Abdullah, Marzbali, Woolley, Bahauddin, & Tilaki, 2014; Hair, Black, Babin, & Anderson, 2010). Discriminant validity can also be evaluated by comparing the AVE with the maximum shared squared variances (MSV) and the average squared shared variances (ASC). Hair et al. (2010) recommended that the criteria of standardized factor loading (standardized regression weights) and AVE should each be ≥ 0.5 and $CR \geq 0.7$. Even so, Hair stated that 0.4 can still be accepted as an adequate factor loading.

The next analysis was conducted by descriptive statistics to examine the teachers' practice profiles in mathematics class for both teaching practices and assessment practices. The presented

descriptive statistics include the mean, standard deviation, range between means, percentage, mean comparison, and effect size. The possible value of the mean is between 1 and 5.

Correlation analysis was conducted to see the consistency between teaching practices and assessment practices. The Spearman correlation was chosen for this analysis because the data for assessment practices were not normally distributed.

C. Findings and Discussion

We divided this section into three sections. The first section is about the scale development study to answer question number 1, the second section is used to examine profiles, and the third section is to see the consistency between related practices.

Findings regarding the first question

Teaching Practice Scale

Analysis of the correlation matrix for the factor analysis of 16 teaching practice items (abbreviated as TP) in mathematics class was performed with the KMO measure of sampling adequacy in which the results were 0.778 and Bartlett's test of sphericity produced a p -value < 0.05 . Thus, the correlation between items of the matrix was suitable for factor analysis.

The analysis was continued with the PCA. There were four factors with eigenvalues of greater than one and also suggested by screen plot. Four factors were determined and followed with the varimax rotation method. Analysis results showed that a few factors were difficult to interpret so that an analysis was done by determining two factors by use of the varimax method. The analysis was then continued by determining the two factors and using the same rotation. The solution was calculated from 38.172% of the total variance, and all of the items ($n = 16$) were used to describe both factors.

The naming of each factor referred to the relationship between items and was then linked to supporting theories (Purnomo, 2016a, 2017b). Factor 1 was linked to relational teaching practices, and factor 2 was linked to instrumental teaching practices. Instrumental teaching practices are related to ways that emphasize more on results than process, not daring to break out of habits, and prioritizes superficial learning. In contrast, relational teaching practices emphasize more on relevancy in the student context, is tolerant with nonconventional ways, and prioritizes meaningful learning.

The internal consistency of the estimated coefficient alpha was 0.743 for relational factor and 0.757 for instrumental factor. The compositions of those factors can be seen in detail in Appendix 1.

After the EFA, CFA by the maximum likelihood (ML) method was conducted with three-time improvement and obtained the output for the final model, which were the value of 53.723; $p = 0.016$; $NC = 1.919$; $RMSEA = 0.053$; $SRMR = 0.043$; $GFI = 0.969$; $AGFI = 0.939$; $NFI = 0.928$; $TLI = 0.941$; $CFI = 0.963$. According to these criteria, the model fit was good.

Each variable in the teaching practice construct had an adequate standardized factor loading value which was in the range of 0.482 and 0.788. Both relational and instrumental factors had a CR of 0.75 which indicated that the variable represented a latent construct. Therefore, convergent validity for the teaching practice construct was adequate, although it had an AVE of < 0.5 . In addition to that, because the AVE was larger than the ASV and MSV, the discriminant validity was adequate.

Analysis of each item showed that the CITC coefficient was greater than 0.3, therefore fulfilling the recommended criteria for item validity. The internal consistency was indicated to

be at an adequate level for each factor with the value of 0.742 for the relational factor and 0.704 for instrumental factor.

Assessment Practice Scale

Analysis of the correlation matrix for the factor analysis of 12 assessment practice items in mathematics class was performed with the KMO measure of sampling adequacy in which the results were 0.810 and Bartlett's test of sphericity produced a p -value < 0.05 . Thus, the correlation between items of the matrix was suitable for factor analysis.

The analysis was continued with the PCA. There were two factors with eigenvalues of greater than one and four factors suggested by screen plot. The varimax rotation was then conducted on these two factors. The solution was calculated from 47.010% of the total variance. All of the items ($n = 12$) were used to describe both factors.

The naming of each factor referred to the relationship between items and was then linked to supporting theories (Delandshere, 2002; Delandshere & Jones, 1999; Purnomo, 2015, 2016b). Factor 1 was linked to summative practices or assessment of learning (AoL), and factor 2 was linked to formative practices or assessment for learning (AfL). AfL assessment practices are informal assessment practices that are continuous and integrated into learning, using ways to obtain feedback for both the teachers to improve their teaching and for the students to know how far and what they are going to do next in learning. In contrast, AoL assessment practices lean toward assessment practices that refer to external standards and are conducted after finishing a lesson unit.

The internal consistency of the estimated coefficient alpha was 0.767 for AoL practice factor and 0.749 for AfL practice factor. The compositions of those factors can be seen in detail in Appendix 2.

After the EFA, CFA by the maximum likelihood (ML) method with Bollen-Stine bootstrapping was conducted with two-time improvement. The final model indicated an acceptable model fit in which the index was within the expected threshold. The obtained indexes were the value of 63.896; $p = 0.001$; $NC = 1.936$; $RMSEA = 0.054$; $SRMR = 0.050$; $GFI = 0.962$; $AGFI = 0.937$; $NFI = 0.916$; $TLI = 0.941$; and $CFI = 0.957$. According to these criteria, the model fit was good.

Each variable in the assessment practice construct had an adequate standardized factor loading value which was in the range of 0.480 and 0.828. The CR results were greater than the recommended threshold of 0.7. The AfL and AoL obtained a value of 0.75 and 0.77. Therefore, convergent validity for the assessment practice construct was adequate, although it had an AVE of < 0.5 . In addition to that, because the AVE was larger than the ASV and MSV, the discriminant validity was adequate.

Analysis of each item showed that the CITC coefficient was greater than 0.3, therefore fulfilling the recommended criteria for item validity. The internal consistency was indicated to be at an adequate level in each factor with 0.733 for AfL factor and 0.731 for AoL factor.

Findings regarding the second question

The second question was: what is the profile of teaching practices in a mathematics class? This question was answered with the analysis results that can be seen in Table 2 (see also Purnomo, 2017b).

Table 2. Practice profile in mathematics class

Dimension	Factor	N	Mean	SD	Mean range of items	Mean comparison	Effect size
PM	Relational	325	3.619	0.593	3.357 – 3.969	$t(324) =$	-0.991
	Instrumental	325	4.324	0.469	4.100 – 4.540	-17.862, $p = 0.000$	
PP	AfL	323	3.347	0.714	3.053 – 3.622	$t(322) =$	-1.404
	AoL	323	4.477	0.449	4.158 – 4.676	-25.225, $p = 0.000$	

According to the analysis summary that can be seen in Table 2, the teaching practice profile of teachers prioritizes instrumental practices rather than relational practices. It can be seen by the mean of relational teaching practices with the value of 3.619 and the mean of instrumental teaching practices with the value of 4.324. This difference was significant as seen by the p-value < 0.05 and had a large size difference in which the effect size had the value of -0.991. A significant and large difference could also be found for the assessment practices. A significant difference was shown by the p-value < 0.05 and a large size difference was shown by the effect size of -1.404. In other words, the assessment practices done by teachers tended to lean toward AoL practices rather than AfL practices in mathematics class.

The research findings of teaching practice profiles of primary school teachers indicated that teachers tend to conduct instrumental teaching practices rather than relational practices. Instrumental teaching practices are very identical to learning based on results rather than process, so learning that is relevant to the context of students is often ignored. These findings strengthen previous studies about the teaching practices of Indonesian mathematics teachers in class (Purnomo, Suryadi, & Darwis, 2016; Wijaya, van den Heuvel-Panhuizen, & Doorman, 2015), in which teacher practices are more dominated by traditional practices. Through class observations, Wijaya found that practices of mathematics teachers, concerning how context-based tasks are taught, leaned more toward a directive teaching approach. Teachers tended to dominate learning activities by delivering information about a problem, about what has to be done and focusing on the mathematics solution without connecting it to the context of the problem. Similar findings can also be found in a study by Purnomo et al. (2016), which indicate that teachers are more dominant as presenters and demonstrators in learning with the use of media and other tools. Meanwhile, students usually see and hear without being involved in the process of using these tools.

Similar to teaching practices, the findings regarding the profile of assessment practices indicate that teachers tended to conduct traditional assessments in mathematics class. The referred assessments are assessments that focus more on formality and accountability aspects rather than focusing on practices that are relevant to the context of learning and the students. The teachers of this sample tended to use tests as a form of assessment and gave marks and scores on the children's worksheets as a form of feedback to them. Several researchers (Hattie & Timperley, 2007; Kasih & Purnomo, 2016; Purnomo, 2014, 2015; Shute, 2007) agreed that, compared to being given marks and scores, feedback in the form of constructive comments are more desired and has a positive impact toward students. Furthermore, Hattie and Timperley (2007) found that useful feedback is giving information to the teachers and students about: where will they go? How far is their position toward the goal that will be obtained? How will I take the next step? It is difficult to achieve through scores and marks as a form of feedback.

Furthermore, teachers in this sample are accustomed to using external assessment standards rather than standards that are appropriate with the students' real conditions. It, of course, separates assessment and learning. As a concrete example, indicators in external tests often are not relevant to the learning done.

A traditional practice that teachers in this sample tend to do is conducting an assessment at the end of a lesson unit. In other words, the assessment information is only based on the results of quizzes, mid-term exams, and national exams. It is indeed not relevant with the purpose of assessment which is to provide information or feedback for both students and teachers to guide learning and teaching to reach a common goal (Kasih & Purnomo, 2016; Purnomo, 2015). That information of course has to be relevant and continuous with the learning and teaching process in class, so it is difficult to maximize the role of assessment as an effort to support learning when assessment practices are still conducted at the end of a lesson unit.

Findings regarding the third question

The third question was: how is the consistency between relevant factors in teaching and assessment practices? Analysis results of the correlation between factors can be found in Table 3 below.

Table 3. The relationship between factors in practice scales

Scale	Factor/Dimension	1	2	3	4
Teaching practice	1. Relational	1	0.091	0.137*	0.059
	2. Instrumental		1	0.131*	0.262**
Assessment practice	3. AfL			1	0.077
	4. AoL				1

** *Correlation is significant at the 0.01 level (2-tailed)*

* *Correlation is significant at the 0.05 level (2-tailed).*

According to the correlation coefficient, as presented in Table 3, the relationship between relevant factors in practice scales in mathematics class was consistent. The most robust relationship was presented by instrumental teaching practice and AoL practice with a correlation coefficient of 0.262 and significant at the alpha level of 1%, followed by the pairing of relational teaching and AfL assessment practice that obtained a correlation coefficient of 0.137 and significant at the level of 5%.

This consistency between relevant factors in practice scales indicate that assessment and teaching practices in class are inseparable parts. On the other hand, a consistent relationship also occurs when teaching practices that focus only on results rather than process guide their assessment practices by referring to external standards that may not be relevant with the content and students during learning.

This consistency is similar to what was stated by Purnomo and colleagues (Kasih & Purnomo, 2016; Purnomo, 2013, 2014, 2015) that the assessment process had become an integrated process in the process of learning. In other words, when assessment practices are used for knowing the extent to which knowledge is taken in after learning, then the teaching by the teacher is more focused on how to guide the students to succeed in the criteria of standardized external assessment. As for the other way around, when teaching practices emphasize more on performance and results, then the assessment practices tend to lean toward the process of seeing the extent to which knowledge is taken in after learning. On the other hand, when teaching practices tend to give room and opportunities for the students to construct their knowledge, then

the assessment practices tend to be used to reflect and guide the learning process. As for the other way around, when the tendency of an assessment process is used to improve and support the learning process, then the teaching practices tend to be relevant with the context and problems of students.

D. Conclusion

This research's purpose was to know the factors that underlie the teaching practices and assessment practices of teachers in mathematics class, to know the overall description of their practices in mathematics class, and to examine the consistency between relevant factors in their practices in mathematics class. First, this research found that there were two factors in each practice scale of teachers in mathematics class. Teaching practices included the relational and instrumental practice factors. Meanwhile, the assessment practices included the practice factors of AfL and AoL. Second, this study found that most practices that were emphasized by teachers went into the direction of traditional practices for both teaching and assessment practices. Lastly, there was a consistency between relevant factors in their practices in mathematics class. It indicated that assessment practices are an integrated part of teaching practices in mathematics class. This study is recommended for researchers, policy makers, and teachers themselves to develop teachers' literacy in regards to assessment and learning. It is of importance as an effort to improve the learning of mathematics. Because, fundamentally, the purpose of assessment is to give effective feedback to students and teachers to improve their practices.

Acknowledgment

We want to thank the Ministry of Research, Technology, and Higher Education of the Republic of Indonesia (Kementerian Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia) for funding this research. We would also like to thank Thahira Hanum Sekarmewangi for improving English of the manuscript.

References

- Abdullah, A., Marzbali, M. H., Woolley, H., Bahauddin, A., & Tilaki, M. J. M. (2014). Territorial functioning and victimisation: Conceptualisation and scale development. *Crime, Law and Social Change*, 61(3), 335–354. Doi:10.1007/s10611-013-9490-6
- Balan, A. (2010). Assessment for learning: A case study in mathematics education. Faculty of Education and Society. Malmö University, Holmbergs, Malmö. Retrieved from <http://www.mah.se/muep>
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824. Doi:10.1016/j.paid.2006.09.018
- Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology*, 99, 288-308. Doi:10.1016/j.jecp.2007.12.002
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, U.K.: Open University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9–21. Retrieved from http://datause.cse.ucla.edu/DOCS/pb_wor_2004.pdf
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148. Doi:10.1002/hrm
- Cherney, A., Povey, J., Head, B., Boreham, P., & Ferguson, M. (2012). What influences the utilisation of educational research by policy-makers and practitioners?: The perspectives of academic educational researchers. *International Journal of Educational Research*, 56, 23-34.

- Doi:10.1016/j.ijer.2012.08.001
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Delandshere, G. (2002). Assessment as inquiry. *Teachers College Record*, 104(7), 1461–1484. Doi:10.1111/1467-9620.00210
- Delandshere, G., & Jones, J. H. (1999). Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision*, 14(3), 216–240. Doi:10.1017/CBO9781107415324.004
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives*, 38(1), 105–123.
- Fabrigar, L. R., MacCallum, R. C., Wegener, D. T., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis (Vol. 7)*.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- James, M., Black, P., Carmichael, P., Conner, C., Dudley, P., Fox, A., Marshall, B. (2006). *Learning how to learn: tools for schools*. New York: Routledge.
- Kasih, P. A., & Purnomo, Y. W. (2016). Peningkatan hasil belajar matematika siswa sekolah dasar melalui pembelajaran berbasis penilaian. *Journal of Research and Advances in Mathematics Education*, 1(1), 69–78.
- Kaur, A., Noman, M., & Awang-Hashim, R. (2018). Exploring and evaluating differentiated assessment practices of in-service teachers for components of differentiation. *Teaching Education*, 62(10), 1–17. Doi:10.1080/10476210.2018.1455084
- Koh, J. H. L., Chai, C. S., & Tsai, C. C. (2010). Examining the technological pedagogical content knowledge of Singapore pre-service teachers with a large-scale survey. *Journal of Computer Assisted Learning*, 26(6), 563–573. Doi:10.1111/j.1365-2729.2010.00372.x
- Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational research in educational practice: Predictors of use. *Canadian Journal of Education*, 37(2), 1-26.
- Ma, X., Millman, R., & Wells, M. (2008). Infusing assessment into mathematics content courses for pre-service elementary school teachers. *Educational Research for Policy and Practice*, 7(3), 165–181. Doi:10.1007/s10671-008-9050-5
- Mansyur, M. (2011). Pengembangan model assessment for learning pada pembelajaran matematika di SMP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(1), 71–91.
- Maulana, R., Opendakker, M.-C., den Brok, P., & Bosker, R. J. (2012). Teacher-student interpersonal behavior in secondary mathematics classes in Indonesia. *International Journal of Science and Mathematics Education*, 10(1), 21–47. Doi:10.1007/s10763-011-9276-1
- Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences (Sixth Edition)*. London, United Kingdom: Taylor & Francis Ltd.
- Purnomo, Y. W. (2013). Keefektifan penilaian formatif terhadap hasil belajar matematika mahasiswa ditinjau dari motivasi belajar. In *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika* (pp. 649–656). Yogyakarta: FPMIPA, Universitas Negeri Yogyakarta.
- Purnomo, Y. W. (2014). Assessment-based learning: Sebuah tinjauan untuk meningkatkan motivasi belajar dan pemahaman matematis. *Sigma*, VI(01), 22–33.
- Purnomo, Y. W. (2015). Pengembangan desain pembelajaran berbasis penilaian dalam pembelajaran matematika. *Cakrawala Pendidikan*, XXXIV(2), 182–191. Doi:10.21831/cp.v2i2.4823
- Purnomo, Y. W. (2016a). *Eksplorasi keyakinan guru tentang matematika, belajar dan mengajar, penilaian dan hubungannya dengan praktik di kelas matematika*. (Unpublished master thesis). Universitas Pendidikan Indonesia, Bandung.
- Purnomo, Y. W. (2016b). Perbaikan instruksional dalam implementasi assessment-based learning di kelas matematika. *Cakrawala Pendidikan*, XXXV(3), 403–411. Doi:10.21831/cp.v35i3.8821
- Purnomo, Y. W. (2017a). A scale for measuring teachers' mathematics-related beliefs: A validity and reliability study. *International Journal of Instruction*, 10(2), 23–38. Doi:10.12973/iji.2017.1022a
- Purnomo, Y. W. (2017b). The complex relationship between teachers' mathematics-related beliefs and their practices in mathematics class. *The New Educational Review*, 47(1), 200–210. Doi:10.15804/ner.2017.47.1.16
- Purnomo, Y. W., Suryadi, D., & Darwis, S. (2016). Examining pre-service elementary school teacher

- beliefs and instructional practices in mathematics class. *International Electronic Journal of Elementary Education*, 8(4), 629–642.
- Setiani, F. (2011). Pengembangan asesmen alternatif dalam pembelajaran matematika dengan pendekatan realistik di sekolah dasar. *Jurnal Penelitian dan Evaluasi Pendidikan*, 15(2), 250–268. Doi:10.21831/pep.v15i2.1096
- Shute, V. J. (2007). *Focus on formative feedback. ETS research report series*. Princeton, NJ: Educational Testing Service.
- Suci, V. W., & Purnomo, Y. W. (2016). Hubungan antara konsepsi penilaian dan kecemasan siswa sekolah dasar di kelas matematika. *Beta: Jurnal Tadris Matematika*, 9(1), 48–60.
- Swan, M. (2006). Designing and using research instruments to describe the beliefs and practices of mathematics teachers. *Research in Education*. Doi:10.7227/RIE.75.5
- Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Teachers' teaching practices and beliefs regarding context-based tasks and their relation with students' difficulties in solving these tasks. *Mathematics Education Research Journal*, 27(4), 637–662. Doi:10.1007/s13394-015-0157-8

Appendix 1

The composition of Teaching Practice Factors in Mathematics Class

	<i>Loading</i>		<i>h²</i>
	1	2	
13. I explain mathematical concepts in order and carefully.	0,711		0,530
11. I prevent students from making mistakes, starting with explaining the content carefully.	0,707		0,446
10. Rules, algorithms, formulas, and methods that I deliver do not deviate from books or curriculum.	0,653		0,506
12. I emphasize on drilling to my students.	0,616		0,380
9. The content that I deliver is per textbooks or student worksheets.	0,611		0,376
15. I use rules, procedures, formulas, or methods that are fast and easy.	0,556		0,409
8. I emphasize how the students follow the rules and procedures correctly.	0,498		0,460
14. Students only use methods that I have taught them.	0,455	-0,345	0,455
16. I teach each topic from the beginning, with the assumption that students are not familiar with the topic yet.	0,402		0,542
2. Students use forms of media and/or tools or technology to develop their mathematical ideas.		0,697	0,446
1. Students or groups of students convey their ideas in front of the class for discussion.		0,650	0,479
4. I integrate other subjects into the subject of mathematics.		0,632	0,375
6. I encourage students to make and discuss mistakes that happen in mathematics class		0,632	0,331
3. I use context and experiences from students to start discussions about the learned topic.		0,590	0,343
7. Students work in pairs or small groups.		0,566	0,496
5. I provide opportunities and guide students to find their own methods.		0,547	0,550
Cronbach's Alpha	0,743	0,757	

Appendix 2.

The composition of Assessment Practice Factors in Mathematics Class

	Loading		h^2
	1	2	
10. The assessment standards that I use refer to established standards.	0,735		0,481
11. The assessment format that I use is for measuring if students are capable of taking in the delivered knowledge.	0,728		0,473
12. I conduct assessments after finishing a lesson unit.	0,705		0,380
9. I give scores or grades on students' worksheets.	0,668		0,536
8. I assess students' work by giving a mark (right or wrong).	0,637		0,498
7. The purpose and criteria of students' successful learning refer to the current curriculum.	0,566		0,393
4. I use rubrics and/or journals in learning activities.		0,718	0,416
5. I give and write descriptive comments on students' worksheets as quickly as possible.		0,700	0,407
1. Students use rubrics and/or journals in their learning.		0,693	0,447
2. The purpose and criteria of success in learning is articulated before discussing the mathematics content.		0,658	0,551
6. I facilitate students to comment on the responses of other students' answers and questions in class.		0,586	0,560
3. Students are given the opportunities to decide their own purpose of learning.		0,585	0,501
Cronbach's Alpha	0,767	0,749	