Title	Contextual factors that affect Singapore secondary school teachers' assessment literacy
Author(s)	Iris Lee, Gavin Fulmer and Kelvin Tan
Source	*39th Annual Conference of the International Association of Educational Assessment (IAEA) on "Educational Assessment 2.0: Technology in Educational Assessment", Tel Aviv, Israel, 20 – 25 October 2013*

Contextual Factors that Affect Singapore Secondary School Teachers' Assessment Literacy

Iris Lee, *Gavin Fulmer and Kelvin Tan*
National Institute of Education, Nanyang Technological University, Singapore

There exist studies such as that of James and Pedder (2006), Brown (2004), Hargreaves, Earl, and Schmidt (2002) which purport to measure teachers' assessment conceptions and practices. However, there has not been any known study that seeks to explore the contextual factors which affect the teachers' assessment literacy. These factors, both socially and contextually value-laden, are likely to influence and/or affect teachers' conceptions of assessment. While teacher literacy regarding assessment principles has been emphasized in preservice training and inservice professional development (Popham, 2009; Stiggins, 2004), such training and development do not occur in vacuums, but are carried out and accepted by people, coming from and embedded within the varied cultures and contexts. As each school is different and performance on assessment literacy may most likely be judged by people operating in these different cultures, what may actually be implemented and operationalized are likely to be subject to peculiarities of these social cultures (Fullan, 2010; Wiggins, 1993). Hence, this study addresses the gap of bridging these missing and yet, essential elements of understanding teachers' conceptions of assessment. A new, 41-item instrument was developed based on Kozma's (2003) framework of micro, meso, and macro-level contextual factors and focus group discussion with teachers, school leaders, and teacher educators. The draft instrument was refined through further teacher interviews to improve legibility and interpretation of the item stems. After pilot testing and analysis with RateFOLD (Luo & Andrich, 1999), a final version was developed with 30 items, and is currently under further study.

Keywords: contextual factors, RateFOLD, micro, meso, macro levels

Email address:

Iris.lee@nie.edu.sg
Gavin.fulmer@nie.edu.sg
Kelvin.tan@nie.edu.sg

Introduction

Teachers' conceptions of assessment (COA) is an area that has been explored. Prior work has focused on the broad purposes of assessment (Brown, 2004) as well as teachers' assessment practices in the classroom (James & Pedder, 2006; Hargreaves, Earl, and Schmidt (2002). However, prior research on teachers' COA have not considered the potential influence of contextual factors that would affect teachers' conceptions and/or literacy. Such contextual factors are embedded within each particular cultural and social context (Fullan, 2010; Wiggins, 1993) so much so that any one social being that exists within the community cannot really deny its possible effects on oneself. A teacher, for instance, is a social being in his/ her community and social cultural context and would hence, be subject to such factors. In other words, if one is to explore teachers' COA, such contextual factors will be worth considering so as to yield a better understanding of COA and implications for research and practice.

Theoretical Framework and Review of Literature

In Kozma's (2003) study on ICT and educational change, he postulated three concentric levels that would affect and mediate changes to teachers' classroom instruction using ICT: the micro, meso and macro levels. Implanting now his three concentric levels to explore the factors that affect the teachers' assessment conceptions, the 3 levels would be teacher level, school level, and national level respectively. The teacher level is the innermost or micro level where most teachers may be concerned with job complexity and autonomy (Deci, Connell & Ryan, 1989) and Rogers' (1995) five factors- relative advantage, the extent to which the new idea, knowledge or innovation is in sync with the teachers' values, experience and needs and how visible the results of the innovation could be observed by others. That is, this level seeks to address the concerns within the teacher's own locus of control. The school organisation would then form the next level. At this meso level, the school leadership, characteristics of the administrators and supervisory support (Tiernery, Farmer & Graen, 1999) would affect the teachers' conceptions about assessment and its purposes. At this level, the school may decide for the teacher and it may be beyond the teacher's jurisdiction to exercise certain decision. The school leadership hence, has an over-riding authority over the teacher's locus of control. The outermost circle or macro level would encompass national policies and international trends. At this level, the situation can be even more complex as the school has to abide by national policies which are often executed at policy levels. If this is the case, how successful a change would be, would then depend on how these various levels are pieced and/or reinforcing one another (Owston, 2007). To make matters even more interesting, any change may likely be 'technically simple and socially complex' (Fullan, 2007, p.84) to begin with. Thus, though Kozma (2003) conceptualized them as concentric circles, these so-called "levels" are not likely to exist in such clear-cut concentric layers. In fact, it would most likely be the optimum bedrock for any change to outlast its challenges when these 3 circles intersect instead of just being concentric (see Figure 1 below). In other words, rhetorically speaking, would addressing these factors that converge at these intersects in turn, be sufficient to facilitate the teachers' assessment conceptions and changes in classroom assessment practices?

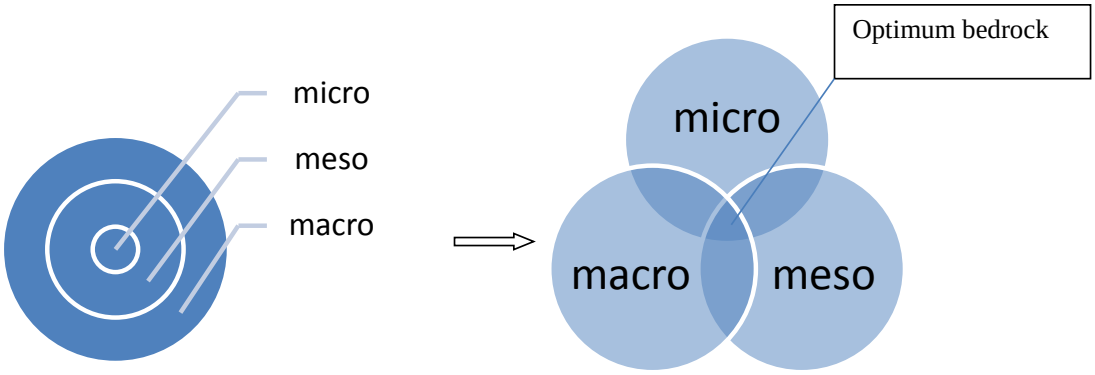


Figure 1: Contextual factors influencing teachers' conceptions of assessment

Background of study

This study is situated in Singapore schools. In 2009, the Ministry of Education (MOE, 2009) announced its pursuit on 'work[ing] with schools to provide schools with training and resources to develop expertise in new teaching and assessment methods, and to ensure that robust assessment frameworks are put in place', 16 prototype schools have taken the lead. '[A] whole school approach' was also recognised as quintessential to building the culture (MOE, 2010) for assessment for learning. Four years on, have recognising such a whole school approach and achieving assessment literacy in schools better support the holistic development of pupils via 'balance[ing] the use of written examinations and alternative modes of assessment' (MOE, 2009)? Popham (2009) asserted that assessment literacy should be part of teachers' professional development and Stiggins (2004) too, emphasised the importance of sound assessment principles. If so, a study on contextual factors on the 3 levels that would examine such a culture may help shed some light. Hence, our main research question for this study is: What are the contextual factors that influence Singapore teachers' assessment literacy and/or COA?

Methods

This study, part of a bigger study, focused on developing and piloting an instrument on contextual factors affecting teachers' assessment literacy, practices and views. The study proceeded in three phases: (1) item creation; (2) item revision; and (3) piloting. In the item creation phase, a focus group discussion (FGD) was first conducted to collect information over what would be the factors that would affect teachers' assessment literacy. The six invited FGD participants held various positions at schools and worked at various educational institutions. They were at least master's degree holders who had done a minimum of an assessment module at master's level too. All 6 of them gave consent for the FGD to be both audio- and video-taped. These tapings were done to facilitate transcription. 44 initial statements or items were developed from the content of the transcripts. They comprise 14 items exploring the micro level, 16 for the meso level, and 14 for the macro level.

In the item revision phase, interviews were conducted with another 6 secondary school teachers. These interviews focused on the legibility and interpretation of the item stems. Interview participants were provided paper copies of the items (without any scale

provided), and invited to talk about the items while also writing on the paper itself to point out areas that were unclear. For instance, two items from the meso level, namely item numbered 14 and 15 were collapsed to form only one item for the intermediate 41-item version:

> Item 14- Teachers need to be trusted for their skills for DnT.
>
> Item 15- Teachers need to be trusted for their skills for SPA.
>
> Collapsed and revised item- Everyone has a voice in the assessment plan (e.g. consensus must be sought from every teacher for assessment plans that includes innovative assessment efforts).

In the piloting phase, this intermediate version was then administered to a sample of 38 teachers. The data were analysed using RateFOLD, a software for generalized unfolding models (Luo & Andrich, 1999). Unlike in classical test theory, where one would just take the summated score of the raw Likert data collected from the survey, item response theory (IRT) allows differences in the extent to which the items and responses reflect the underlying trait, yielding more accurate estimates of this latent trait of the sample. Taking summation of the raw scores could not provide such information. Furthermore, when items have the potential of being ambivalent, that is, when a person's level of agreement could be due to their being higher or lower on the trait than the item itself, the unfolding model yields even more accurate estimates of each person's latent trait. Therefore, a unidimensional unfolding model was chosen and estimated using RateFOLD so as to construct an instrument that would measure contextual factors of assessment literacy/ COA. In the arena of attitudinal measurement, the fit indices would be the appropriate check on the instrument reliability than that of internal consistency through Cronbach's alpha, as in studies for achievement scores.

Preliminary Results and analysis

Focus group discussions. The video recording from the focus group discussion was transcribed and this transcript was analysed. Sentences and phrases were coded and further collapsed to fit into the 3 levels. Drawing on these levels within the discussion, an initial 44-item instrument was developed (see Appendix A for sample items). This initial instrument was checked for legibility and interpretation of the items by another 6 secondary school teachers. The improved version where 3 items were either deleted or reworded into the other items was then piloted.

Pilot study. Teachers taking part in an intensive summer Master's degree course were recruited for the pilot study. A sample of 38 teachers submitted completed questionnaires. The data was analysed via the RateFOLD program. Initial correlational analysis (see Appendix B) was done after keying the raw data into excel spreadsheet. If this dataset were to yield all positive correlations, it would be beneficial to use software that would support the cumulative model for further analysis, since the positive correlations would suggest the Guttman's (1950) cumulative structure. However, the ambivalent nature of this data, as denoted by the mixture of negative and positive correlations in Appendix B found RateFOLD to be a suitable program. Further qualitative item-by-item checks for all of the 41-items found

that the program was assigning the signs (either positive or negative to the central notion of each level) correctly and a probability of p=0.086 >0.01 was obtained for the 41-item survey (see item map in Appendix C). This means that the data fits the model and this fit-statistic is equivalent to that of obtaining a good score on reliability as in the case of an achievement test. As the original plan was to yield a total of 30-item from this phase before further work, selection was made via selecting the items with the highest probability and smallest chi (see Appendix D). Hence, a total of 11 items were deleted. The instrument is now ready for further use.

Conclusion

Previous research has paid relatively little attention to the contextual factors that may affect teachers' assessment conceptions, views, or practices. This is despite the importance of such factors in affecting teachers' views and work (Kozma, 2003). Addressing this gap requires research work on possible contextual factors and ways to measure teachers' experience of them. To address this gap, the present study developed and piloted an instrument that measures contextual factors at three levels: micro, meso, and macro. Preliminary analysis indeed yielded a 30-item survey which serves to support the existence of contextual factors. These contextual factors do exist in 3 levels and further collection of data may shed more understanding on the confluence of such factors.

Formatted: Font: Bold

Acknowledgement

The authors would like to thank Professor Luo GZ for the use of the RateFOLD (Luo & Andrich, 1999) program used in this paper.

References

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.

Deci, E.L., Connell, J.P., & Ryan, R.M. (1989). Self-determination in a work organisation. *Journal of Applied Psychology, 74*, 580-590.

Fullan, M. (2007). *The new meaning of educational change* (4th ed.). New York: Teachers college, Columbia

Fullan, M. (2010). *All Systems Go: The Change Imperative for Whole System Reform.* CA: Corwin.

Guttman, L. (1950). The basis for scalogram analysis. In S.A. Stouffer, et al. (Eds.), *Measurement and prediction* (pp. 60-90). New York: Wiley.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69-95.

James, M., & Pedder, D. (2006). Beyond method: assessment and learning practices and values. *The Curriculum Journal, 17*(2), 109-138.

Kozma, R.B. (2003). *Technology, innovation, and educational change: a global perspective*. Eugene, OR: International society for technology in education.

Luo, G., & Andrich, D. (1999). *RateFOLD: Unfolding analysis program for polytomous responses*. Singapore: Nanyang Technological University of Singapore.

Ministry of Education (2009). Report of the Primary Education Review and Implementation Committee. Singapore: Ministry of Education.

Ministry of Education. (Jul, 2010). Press Release. PERI Holistic Assessment Seminar 2010 — 16 Schools Share Holistic Assessment Practices and Resources. Retrieved August 12, 2011 from http://www.moe.gov.sg/media/press/2010/07/peri-holistic-assessment-seminar.php

Owston, R. (2007). Contextual factors that sustain innovative pedagogical practice using technology: An international study. *Journal of Education Change, 8*(1), 61-77.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice, 48*, 4-11.

Rogers, E.M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

Stiggins, R. (2004). New assessment beliefs for a new mission. *Phi Delta Kappan, 86*, 22-27.

Tierney, P., Farmer, S. M., & Graen, G.B. (1999). An examination of leadership and employee creativity: the relevance of traits and relationships. *Personnel Psychology, 52*, 591-620.

Wiggins, G. (1993). Assessment: Authenticity, context and validity. *Phi Delta Kappan, 75*(3), 200.

Appendix A- Sample items for initial 44-item survey

Micro level- self as locus of control

(Training, relational, accountability to parents, stakeholders)

1. I feel competent to design assessment needed for my class.
2. I am competent at designing assessment needed for my class.
3. I need more knowledge in order to design good assessment for my class.

Meso level- school level as locus of control

(PLC, assm consultancy, resources, wkload, support fm HOD, SLs, time for planning, accountability to parents, stakeholders)

1. My school culture is supportive of my needs for formative assessment.
2. My colleagues are supportive of one another as we explore new assessment.
3. I am not afraid of being ranked poorly in my school if my students' grades are not up to their expectation due to the new assessment strategies that I take.

Macro level- national policy as locus of control

(accountability to parents, stakeholders, power disseminated)

1. I wish teachers were allowed some risk to fail at new assessment strategies.
2. There exists only calculated risk for failure even with assessment issues.
3. Policy must support innovative assessment trials at school.

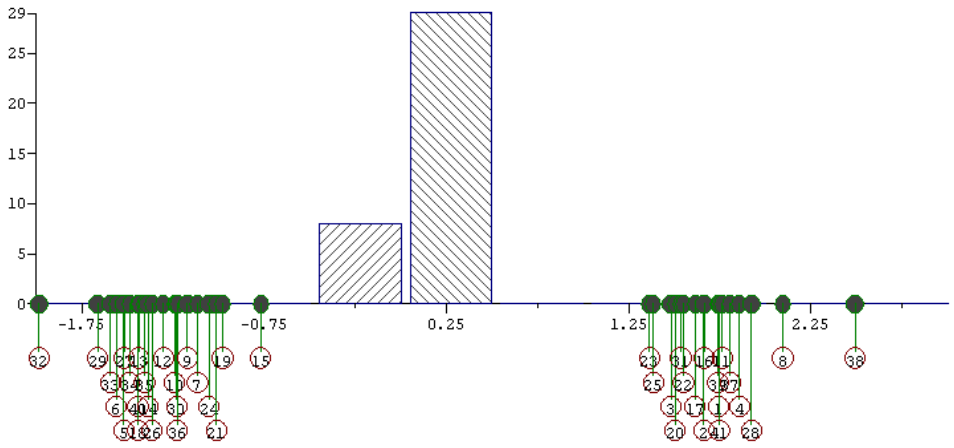
Appendix B- Sample of correlational analysis of the 41-item survey

	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00											
2	0.13	1.00										
3	0.03	0.30	1.00									
4	0.64	0.14	(0.00)	1.00								
5	0.23	0.49	0.24	0.30	1.00							
6	0.13	(0.10)	(0.21)	0.21	(0.23)	1.00						
7	(0.05)	0.08	0.23	(0.13)	0.16	0.34	1.00					
8	0.17	0.07	0.08	0.28	0.27	0.02	(0.06)	1.00				
9	0.04	0.27	0.11	0.19	0.34	0.21	0.34	0.16	1.00			
10	0.03	0.16	0.15	(0.09)	0.32	0.06	0.41	0.17	0.40	1.00		
11	0.13	0.13	(0.38)	0.09	0.23	0.24	0.14	0.22	0.21	0.18	1.00	

12	0.24	(0.04)	0.05	0.10	0.09	0.25	0.21	0.16	0.30	0.32	0.56	1.00

( )-Bracket denotes negative figures

Appendix C- Item and person location map of 41-item survey



Appendix D- Selection of 30-item from the 41-item survey (highlighted ones are discarded)

1. Df= 12, ChiSQ= 8.2360, Prob= 0.7664
2. Df= 12, ChiSQ= 10.9381, Prob= 0.5342
3. Df= 12, ChiSQ= 10.5574, Prob= 0.5672
4. Df= 12, ChiSQ= 6.0749, Prob= 0.9123
5. Df= 12, ChiSQ= 7.1380, Prob= 0.8484
6. Df= 12, ChiSQ= 4.4370, Prob= 0.9742
7. Df= 12, ChiSQ= 6.3989, Prob= 0.8947
8. Df= 12, ChiSQ= 5.5855, Prob= 0.9355
9. Df= 12, ChiSQ= 9.3136, Prob= 0.6759
10. Df= 12, ChiSQ= 10.0449, Prob= 0.6120
11. Df= 12, ChiSQ= 6.4303, Prob= 0.8929
12. Df= 12, ChiSQ= 5.6703, Prob= 0.9318
13. Df= 12, ChiSQ= 13.4205, Prob= 0.3392
14. Df= 12, ChiSQ= 5.0512, Prob= 0.9562
15. Df= 12, ChiSQ= 7.6393, Prob= 0.8126
16. Df= 12, ChiSQ= 16.3747, Prob= 0.1747
17. Df= 12, ChiSQ= 11.8726, Prob= 0.4560
18. Df= 12, ChiSQ= 6.4792, Prob= 0.8900

19. Df= 12, ChiSQ= 5.0163, Prob= 0.9574
20. Df= 12, ChiSQ= 7.5394, Prob= 0.8200
21. Df= 12, ChiSQ= 3.8537, Prob= 0.9859
22. Df= 12, ChiSQ= 3.9581, Prob= 0.9842
23. Df= 12, ChiSQ= 6.5842, Prob= 0.8838
24. Df= 12, ChiSQ= 5.7360, Prob= 0.9288
25. Df= 12, ChiSQ= 21.7215, Prob= 0.0408
26. Df= 12, ChiSQ= 8.5171, Prob= 0.7435
27. Df= 12, ChiSQ= 8.2839, Prob= 0.7626
28. Df= 12, ChiSQ= 11.1986, Prob= 0.5120
29. Df= 12, ChiSQ= 22.1405, Prob= 0.0360
30. Df= 12, ChiSQ= 4.9947, Prob= 0.9582
31. Df= 12, ChiSQ= 7.7388, Prob= 0.8052
32. Df= 12, ChiSQ= 16.1332, Prob= 0.1852
33. Df= 12, ChiSQ= 6.8280, Prob= 0.8688
34. Df= 12, ChiSQ= 6.7274, Prob= 0.8751
35. Df= 12, ChiSQ= 9.2232, Prob= 0.6838
36. Df= 12, ChiSQ= 5.6099, Prob= 0.9345
37. Df= 12, ChiSQ= 5.0668, Prob= 0.9557
38. 9Df= 12, ChiSQ= 11.2308, Prob= 0.5093
39. Df= 12, ChiSQ= 12.5012, Prob= 0.4063
40. Df= 12, ChiSQ= 8.4628, Prob= 0.7480
41. Df= 12, ChiSQ= 8.3465, Prob= 0.7575