STUDENTS' PERCEPTIONS OF ASSESSMENT PROCESS: QUESTIONNAIRE DEVELOPMENT AND VALIDATION

Bruce G. Waldrip
University of Southern Queensland
Australia
Darrell L. Fisher
Curtin University of Technology
Australia
Jeffrey P. Dorman
Australian Catholic University
Australia

ABSTRACT

Research aimed at developing and validating an instrument to assess middle school students' perceptions of assessment was conducted. Following a review of literature, a tentative 6-scale instrument of 48 items was trialled with a sample of 320 students in 7 Australian schools. Based on internal consistency reliability data and exploratory factor analysis, refinement decisions resulted in a 5-scale instrument called the Student Perceptions of Assessment Questionnaire (SPAQ). The scales of the SPAQ are Congruence with Planned Learning, Authenticity, Student Consultation, Transparency, and Diversity. A sample of 3,098 students in 150 classrooms was used to validate the final SPAQ.

INTRODUCTION

Despite the growth in emancipatory conceptualisations of classrooms that embrace a constructivist epistemology, little contemporary evidence exists to support the view that students are genuinely involved in decision-making about their assessment tasks. That is, forms of assessment and specific assessment tasks employed in schools are overwhelmingly decided by teachers and administrators. Furthermore, even though reports like *The Status and Quality of Teaching and Learning in Australia* (Goodrum, Hackling, & Rennie, 2001) have asserted that assessment is a key component of the teaching and learning process, teachers tend to utilise a very narrow range of assessment strategies on which to base feedback to parents and students. In practice, there is little evidence that teachers actually use diagnostic or formative assessment strategies to inform planning and teaching (Radnor, 1996). This could be due to teachers feeling that they need to 'sacrifice learning with understanding for the goal of drilling students in the things for which they will be held accountable' (Hobden, 1998, p. 221).

Historically, teachers have received substantial levels of advice on assessment practices. Harlen (1998) advises teachers that both oral and written questions should be used in assessing student's learning. The inclusion of alternative assessment strategies, such as teacher observation, personal communication, and student performances, demonstrations, and portfolios, have been offered by experts as having greater usefulness for evaluating students and informing classroom instruction (Brookhart, 1999; Stiggins, 1994). Based on research with teachers, Barksdale-Ladd and Thomas (2000) identified five best practices in assessment:

- providing feedback to help students improve their learning;
- conceptualising assessment as part of a student's work, which can go into a working portfolio;
- providing flexibility so that assessment does not dominate the curriculum;
- ensuring that assessment informs instruction to help teachers improve their teaching, thereby ensuring student learning; and
- using more than one measuring stick to assess students' learning.

Reynolds, Doran, Allers, and Agruso (1995) argued that for effective learning to occur, congruence must exist between instruction, assessment and outcomes.

Few textbooks on classroom teaching and assessment suggest a substantive role for students in developing assessment tasks. This position is historically and culturally based and is rooted in an outdated "assembly-line" view of learning in which recitation of facts is highly prized. In today's information age, jobs are increasingly demanding higher levels of literacy skill and critical thinking and these demands require students to actively engage and monitor their learning rather than passively receive knowledge. This requires a fundamental review of how teachers involve students in assessment tasks (Rogoff, 2001).

An effective assessment process should involve a two-way communication system between teachers and their students. Historically, teachers have used testing instruments to transmit to the student and their parents what is really important for the student to know and do. While this reporting tends to be in the form of a grade, the form and design of assessment can send subtle messages on what is important. There has been a substantial amount of research into

types of assessment but very little research into students' perceptions of assessment (see e.g., Black & Wiliam, 1998; Crooks, 1998; Plake, 1993, Popham, 1997).

In one of the few studies conducted on students' perceptions of assessment, an American sample of 174 students in Years 4 to 12 responded to a specially-designed questionnaire (Schaffner, Bury, Stock, Cho, Boney, & Hamilton, 2000). This research, which also elicited teachers' self-reported perceptions of competence in the design and implementation of assessment tasks, found that teachers were not asking students about what should be included in assessment tasks. By including students in the teaching – testing – grading cycle, the validity of the assessment processes can be enhanced and invalid assessment instruments that result in very high failure rates can be avoided (see e.g., Steinberg, 2000).

THE PRESENT STUDY

Aim

The aim of the present study was to develop and validate an instrument to assess students' perceptions of assessment tasks for use with middle school students.

Sample

In the initial trial validation, a sample of 320 middle school students from Western Australia and Queensland responded to the tentative form of the Students' Perceptions of Assessment Questionnaire (SPAQ) described below. The final form was administered to 3,098 students in 150 classrooms.

Methodological approach to development of the SPAQ

Fraser (1986) and Hase and Goldberg (1967) identified four approaches to the development of instruments: intuitive-rational, intuitive-theoretical, factor analytic and empirical group discriminative. While intuitive-rational and intuitive-theoretical scales rely on the nomination of items to tentative scales prior to questionnaire administration, factor analytic scales employ factor analysis to group items solely on the responses of a sample of the target population being investigated. Empirical group discriminative scales also require test administration prior to scale formation but they are align with an external criterion by selecting items that maximise discrimination between groups of respondents. The SPAQ was developed using an intuitive rational approach to instrument design and validation.

The validity of intuitive-rational scales rests partly on the subjective opinions of the investigators and other experts. There are three procedural steps to intuitive-rational scale development: (1) identification of salient dimensions, (2) writing sets of test items that are linked conceptually with each salient dimension, and (3) field testing the questionnaire. Identifying salient dimensions usually involves a review of literature and utilises the researchers' academic expertise. Writing test items utilises the subjective opinions of researchers with scale development knowledge. Field testing involves administration of the questionnaire to a sample of the target population and studying the internal consistency (usually employing the Cronbach α coefficient as an index) and discriminant validity (using the mean correlation of a scale with the remaining scales as a convenient index) of each scale. Factor analysis can also be used to assist with scale refinement.

DEVELOPMENT OF A TENTATIVE FORM OF THE STUDENTS' PERCEPTIONS OF ASSESSMENT QUESTIONNAIRE (SPAQ)

Identification of salient dimensions

As indicated above, the first step in the intuitive-rational approach is the identification of salient dimensions. An extensive literature review was conducted with an analysis of over 64 policy and research papers on exemplary and effective assessment practices, characteristics, principles, and techniques. For example, McMillan (2000) identified authenticity, feedback opportunities, validity, fairness, ethical, efficient, feasible and utilising multiple methods as important characteristics of assessment in schools. The North West Regional Educational Laboratory's (1995) research synthesis on effective schooling practices lists many characteristics of quality assessment in schools including: reviewing assessment instruments and methods for cultural and other bias, aligning assessments of student performance with the written curriculum and actual instruction, teaching students to evaluate their own work through peer and self-assessment.

Dietel, Herman and Knuth (1991) noted several important characteristics of good assessment. Such assessment: provides accurate estimates of student performance, is consistent and reliable, tests the full range of knowledge and skills, and is free of extraneous factors that confuse students. Furthermore, they argue that good assessment involves

students in setting goals and criteria for assessment and performing tasks that measure meaningful instructional activities. Such activities should be contextualised in real-world situations. Stern and Algren's (2002) review of assessment in science curriculum materials employed three assessment criteria: the extent to which assessment tasks align with the goals of the materials, the extent to which the items focus on student understanding, and the extent to which assessment informs instruction.

Another important source for the present study was the *Perceptions of Assessment* questionnaire developed by Schaffner et al. (2000). This 55-item questionnaire asked students to respond on "how you feel about the way your teacher finds out how much you have learned". Two scales of this questionnaire are particularly relevant to the present study: Fairness Issues, and Student Input into Grading. In Australia, the *Student Learning Preferences* questionnaire (Gough, Waldrip, Tytler, Beeson, & Sharpley, 2002) was reviewed.

Based on this review of policy and research papers, the following tentative dimensions were deemed salient to the present instrument: Congruence with Planned Learning, Diverse Methods, Authenticity, Student Consultation, Transparency, and Accommodation of Student Diversity. Table 1 shows these dimensions and their common sense definitions. These definitions were based on the literature review described above.

Table 1
Description of the SPAQ (Tentative Form)

Dimension	*Description*
Congruence with Planned Learning	The extent to which assessment tasks align with the goals, objectives and activities of the learning program.
Diverse Methods	The extent to which multiple, varied assessment tasks are employed.
Authenticity	The extent to which assessment tasks feature real life situations that are relevant to the learner.
Student Consultation	The extent to which students are consulted and informed about the forms of assessment tasks being employed.
Transparency	The extent to which the purposes and forms of assessment tasks are well-defined and clear to the learner.
Accommodation of Student Diversity	The extent to which all students have an equal chance at completing assessment tasks.

Writing of test items

The second step in the intuitive-rational approach to scale development requires sets of items that are conceptually linked with each salient dimension to be written. Because of the need to limit the length of the instrument, it was decided to write a set of 11 items for each dimension and subject these to measurement scrutiny with a goal of having eight per scale in the tentative instrument. Accordingly, a pool of 66 items was checked for faults and ambiguities by a group of academics with expertise in educational and psychological measurement and school assessment. Particular attention was paid to the face validity and the scale allocation of each item. These items employed a 4-point Likert scale response: Almost Never, Sometimes, Often, Almost Always. The result of this review process was a 48-item instrument with six scales. Each scale has eight items.

Field testing

In line with the third step in the intuitive-rational approach to scale development, the SPAQ (tentative form) was field tested. The sample of 320 Australian middle school students describe above responded to the SPAQ (tentative form). The internal consistency reliability (Cronbach Coefficient α) of each scale and the discriminant validity (mean correlation of each scale with the remaining scales) were computed. Table 2 shows these values.

Apart from the Diverse Methods scale, all scales had satisfactory internal consistency reliability (see Table 2). Discriminant validity data indicated that the scales do overlap. And exploratory factor analysis with varimax rotation revealed low factor loading for items of the Diverse Methods scale.

Table 2
Validation Data for SPAQ (Tentative Form)

Dimension	Internal Consistency Reliability (Coefficient α)	Discriminant Validity (Mean Correlation)
Congruence with Planned Learning	.78	.45
Diverse Methods	.48	.34
Authenticity	.76	.51
Student Consultation	.69	.55
Transparency	.88	.55
Accommodation of Student Diversity	.68	.44

Refinement decisions

Based on the above data, and a review of the scales, the following decisions were implemented. First, the Diverse Methods scale was deleted. This decision was based on two grounds: poor internal consistency reliability and conceptual overlap with the Accommodation of Student Diversity scale. Second, the Accommodation of Student Diversity scale was renamed Diversity. Third, several items were modified to enhance their face validity. Finally, it was decided to reduce the number of items in each scale to six. This decision was based on the data which indicated no appreciable loss of internal consistency reliability for the five scales. Additionally, shorter scales enhance the overall economy of administration of the instrument. A copy of this final form of the SPAQ is in the Appendix.

Table 3
Descriptive Information for the SPAQ (Final Form)

Scale Name	Description	Internal Consist. (Coeff't α)	Discrim. Validity (Mean Corr.
Congruence with Planned Learning	The extent to which assessment tasks align with the goals, objectives and activities of the learning program.	.77	.45
Authenticity	The extent to which assessment tasks feature real life situations that are relevant to the learner.	.72	.50
Student Consultation	The extent to which students are consulted and informed about the forms of assessment tasks being employed.	.68	.55
Transparency	The extent to which the purposes and forms of assessment tasks are well-defined and clear to the learner.	.86	.58
Diversity	The extent to which all students have an equal chance at completing assessment tasks.	.74	.45

Table 4
Factor Loadings for 30 Items of the SPAQ (Final Form)

Item	Factor Loading				
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
CPL1	.70				
CPL2	.54				
CPL3	.71				
CPL4	.60				
CPL5	.57				
CPL6	.70				
A1	.56				
A2	.33				
A3	.32				
A4	.63				
A5	.70				
A6	.44				
SC1					
SC2					
SC3			.53		
SC4			.70		
SC5			.57		
SC6			.57		
T1				.47	
T2				.51	
T3				.53	
T4				.71	
T5				.70	
T6				.58	
D1					
D2					.41
D3					.67
D4					.67
D5					.78
D6					.62

Note. Factor loadings below .30 have been omitted. CPL: Congruence with Planned Learning, A: Authenticity, SC: Student Consultation, T: Transparency, S: Diversity

VALIDATION OF THE FINAL FORM OF THE SPAQ

The final form of the SPAQ consists of five 6-item scales assessing Congruence with Planned Learning, Authenticity, Student Consultation, Transparency, and Diversity. Table 3 shows internal consistency reliability and discriminant validity data for the final form of the SPAQ based on the sample of 320 Australian students. All scales have satisfactory internal consistency reliability. Discriminant validity data indicate that the scales overlap. However, their conceptual distinctiveness warrants retention of all five scales.

Exploratory factor analysis with a varimax rotation revealed five factors accounting for 51.5% of variance in scores. Table 4 shows that 27 of the 30 items had loadings in excess of .30 on their *a priori* scales. This indicates a sound instrument structure.

The final version was administered to 3,098 students in 150 classrooms. Table 5 shows that the final version had acceptable reliability. The reliability of the SPAQ was evaluated by subjecting the data to item analysis and the internal consistency/reliability (Cronbach alpha reliability) for the factors are shown in Table 5. The data in the table shows that for this sample, the alpha reliability ranged from 0.62 to 0.82 suggesting that each SPAQ scale have acceptable reliability, especially for scales containing a relatively small number of items. The mean partial correlation of a scale with other scales was used as a convenient measure of the discriminant validity of the SPAQ. The mean partial correlation value of a scale with other scales is reported in table 5. The mean correlations ranged from 0.38 to 0.46 indicating that the SPAQ measures distinct, although somewhat overlapping aspects of dimensions of assessment. Different teacher tend to use a different selection of assessment tasks. For an instrument to be valid and reliable, it should detect these differences. Eta^2 is a statistical concept that guides researchers on these differences. The eta^2 data was computed using a one-way ANOVA with class membership as the main effect .The instrument recorded statistically significant eta^2 values range 0.14 – 0.20, $p = 0.000$) for all scales suggesting its suitability in locating differences students' perceptions in different classes.

Table 5
Validation Data for SPAQ (Final Form)

Dimension	Internal Consistency Reliability (Coefficient α)	Discriminant Validity (Mean Correlation)	ANOVA Results (Eta^2)	Scale Mean (range 1-4)
Congruence with Planned Learning	.72	.38	0.15	3.07
Authenticity	.82	.43	0.20	2.33
Student Consultation	.72	.44	0.20	2.25
Transparency	.82	.46	0.14	3.04
Diversity	.62	.44	0.15	2.57

The average scale item mean values in each of the five scales are shown in table 5. The values ranged from 2.33 to 3.07. The average scale item mean values were greatest for Congruence with Planned Learning and Transparency while it was lowest for Authenticity, Student Consultation on Assessment and Diversity. The higher average scale item mean values for Congruence with Planned Learning and Transparency scales suggest that in general students perceived that often the assessment covered what they learned in their classes and there was often transparency in their assessment. However, the lower value for Student Consultation on Assessment suggests a low level consultation with students. The data for the remaining three scales, Authenticity and use of Diversity in Assessment scales suggests a week link between assessment and application of knowledge to daily life, and, only sometimes their assessment is catered for students' diversity. The overall analysis of students' perception data suggests a scope for improvement. It is possible that a change is required at the part of teachers and persons involves in assessment.

CONCLUSION

This paper has reported the initial development and validation of an instrument to assess middle school students' perceptions of assessment. As indicated earlier in this paper, students have been left out of assessment deliberations in spectacular fashion for many years and the availability of a validated instrument could facilitate student involvement in the process. Following an intuitive-rational approach to instrument development, a

tentative 6-scale instrument was field tested and subsequently a 5-scale structure was accepted. The development of this instrument, the Students' Perceptions of Assessment Questionnaire (SPAQ) was the initial stage of a larger study.

This research is different in that it examines their perceptions of the process of assessment. In fact, much previous research examines the impact of student peer-assessment on learning. How teachers respond to these perceptions will be varied. Some teachers will accept them as is and try to teach students in somewhat mechanistic ways that will reinforce these perceptions. Other teachers will endeavours to make changes to how students perceive the assessment process. As the U.N. Convention on Children's Rights infer, these teachers will engage students in the assessment process and will give these students a voice that will illuminate students' perceptions of assessment. There will be students who believe that the assessment process is detached from them and that the student should have no input into the process. One could then examine whether involving the students in the process does change their views.

REFERENCES

Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: teachers and parents speak out. Journal *of Teacher Education, 51,* 384-397.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5* (1), 7-74.

Brookhart, S.M. (1999). The art and science of classroom assessment: The missing part of pedagogy. *ASHE-ERIC Higher Education Report. 27* (1).

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58,* 438-481.

Dietel, R. J., Herman, J. L. & Knuth, R. A. (1991). *What does research say about assessment?* Retrieved June 5, 2004, from http://www.ncrel.org/sdrs/areas/stw_esys/4assess.htm

Fraser, B. J. (1986). *Classroom environment*. London: Croom Helm.

Goodrum, D., Hackling, M., & Rennie, L. (2001). *The status and quality of teaching and learning in Australian schools.* Department of Education, Training and Youth Affairs: Canberra.

Gough, A., Waldrip, B.G., Tytler, R., Beeson,G. & Sharpley, B. (2002, April). *Improving Effective Science Teaching and Learning in Australian Schools.* A paper presented at the Annual Meeting of the National Association of Research in Science Teaching, New Orleans.

Hase, H. D., & Goldberg, L. G. (1967). Comparative validity of different strategies of constructing personality inventory scales. *Psychological Bulletin, 67*, 231-248.

Hobden, P. (1998). The role of routine problems in science teaching. In B. Fraser & K. Tobin (Eds.), *International handbook of science education* (pp. 219-232). Dordrecht, The Netherlands: Kluwer.

McMillan, J. A. (2000). *Basic assessment concepts for teachers and school administrators.* ERIC/AE Digest. (ERIC Document Reproduction Service No. ED447201)

North West Regional Educational Laboratory. (1995). *Classroom characteristics and practices.* Retrieved June 5, 2004, from http://www.nwrel.org/scpd/esp/esp95_1.html

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6*, 21-27.

Popham, W. J. (1997). Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13.

Radnor, H. (1996). *Evaluation of key stage3 assessment in 1995 and 1996* (research report). Exeter: University of Exeter.

Reynolds, D.S., Doran, R.L., Allers, R.H., & Agruso, S.A. (1995). *Alternative assessment in science: A teacher's guide.* Buffalo, NY: University of Buffalo.

Rogoff, B. (2001, September 14). Student assessment for the information age. *The Chronicle of Higher Education, 48*(3), p. B17.

Schaffner, M., Burry-Stock, J.A., Cho, G., Boney, T., & Hamilton, G. (2000, April). *What do kids think when their teachers grade*? Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA).

Steinberg, J. (2000, December 22). Student failure causes states to retool testing programs. *The New York Times,* p. A1.

Stern, L., & Ahlgren, A. (2002) Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching, 39,* 889-910.

Stiggins, R. (1994). *Student-centered classroom assessment*. Ontario: Macmillan College Publishing Co.

Acknowledgments

The research described in this paper was supported in part by an Australian Research Council Discovery Grant DP0451981.

Appendix

Students' Perceptions of Assessment Questionnaire (SPAQ)

1. Questions in science tests what I know.
2. My science assignments/tests examines what I do in class.
3. My assignments/tests are about what I have done in class.
4. How I am assessed is like what I do in class.
5. How I am assessed is similar to what I do in class.
6. I am assessed on what the teacher has taught me.
7. I am asked to apply my learning to real life situations.
8. My science assessment tasks are useful in everyday things.
9. I find science assessment tasks are relevant to what I do outside of school.
10. Assessment in science tests my ability to apply what I know to real-life problems.
11. Assessment in science examines my ability to answer every day questions
12. I can show others that my learning has helped me do things.
13. In science I am asked about the types of assessment that are used.
14. I am aware how my assessment will be marked.
15. I can select how I will be assessed in science.
16. I have helped the class develop rules for assessment in science.
17. My teacher has explained to me how each type of assessment is to be used.
18. I have a say in how I will be assessed in science.
19. I understand what is needed in all science assessment tasks.
20. I know what is needed to successfully complete a science assessment task.
21. I am told in advance when I am being assessed.
22. I am told in advance on what I am being assessed.
23. I am clear about what my teacher wants in my assessment tasks.
24. I know how a particular assessment task will be marked.
25. I have as much chance as any other student at completing assessment tasks
26. I complete assessment tasks at my own speed.
27. I am given a choice of assessment tasks.
28. I am given assessment tasks that suit my ability.
29. When I am confused about an assessment task, I am given another way to answer it.
30. When there are different ways I can complete the assessment.

Scale Allocations:
Congruence with Planned Learning: 1-6
Authenticity: 7-12
Student Consultation: 13-18
Transparency: 19-24
Diversity: 25-30