

# Using the student perceptions of assessment questionnaire (SPAQ) to develop an assessment typology for science classes

## Utilización del cuestionario de valoración de percepciones estudiantiles (SPAQ) para desarrollar una tipología de evaluación para clases de ciencias

JEFFREY P. DORMAN <sup>1</sup>, BRUCE G. WALDRIP <sup>2</sup>, DARRELL L. FISHER <sup>3</sup>

<sup>1</sup> School of Education, Australian Catholic University, P.O. Box 456, Virginia 4014

<sup>2</sup> University of Southern Queensland

<sup>3</sup> University of Southern Queensland, Australia

J. Dorman@mcauley.acu.edu.au

### Abstract

*This article reports research conducted on students' perceptions of assessment in science classes in Queensland and Western Australia. A specially developed instrument, the Student Perceptions of Assessment Questionnaire (SPAQ) which assesses Congruence with Planned Learning, Authenticity, Student Consultation, Transparency, and Diversity was used to collect data from 3,055 students. Hierarchical cluster analysis resulted in a four cluster solution being accepted. While one cluster of 799 students held positive perceptions of assessment, another cluster of 640 students held negative views. The SPAQ allows for a greater focus on classroom-based perceptions of assessment rather than crude external accountability measures that decontextualise classroom assessment.*

**Key words:** *perceptions of assessment, classroom typology, science, teaching*

### Resumen

*Este artículo presenta una investigación sobre las evaluaciones a los estudiantes en clases de ciencias en Queensland y Australia Occidental. Para esto en el grupo de 3055 estudiantes se utilizó el cuestionario de valoración de percepciones estudiantiles (SPAQ) que evalúa la congruencia del aprendizaje planeado, la autenticidad, consultas de los estudiantes, la transparencia, y la diversidad de evaluación. En este caso se aplicó un método especial, que mostró, que mientras un grupo de 799 estudiantes tuvieron impresión positiva sobre la evaluación, otro grupo de 640 estudiantes mostró sus impresiones negativas. El SPAQ permite un enfoque más preciso sobre estas impresiones de los estudiantes en el aula, comparadas con herramientas tradicionales.*

**Palabras clave:** *impresiones de los estudiantes, evaluación, tipología, ciencias, enseñanza.*

### INTRODUCTION

One observation of contemporary schools is that forms of assessment and specific assessment tasks employed in schools are overwhelmingly decided by teachers and administrators. Indeed, there is little contemporary evidence to support the view that students are genuinely involved in deci-

sion-making about their assessment tasks. Furthermore, teachers utilise a very narrow range of assessment strategies. The purpose of the present study was to use a newly developed instrument, the Students' Perceptions of Assessment Questionnaire (SPAQ) to study the characteristics of relatively homogeneous groups of students base on their perceptions of assessment tasks. Before providing details of this study, salient literature on students' perceptions of assessment is reviewed.

### Students' perceptions of assessment

Over a significant time period, teachers have received substantial levels of advice on assessment practices. Harlen (1998) advised 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 (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.

In the USA, assessment of student learning has become highly bureaucratised with high-stakes testing procedures evident in most states.

In Australia, a similar trend is developing with benchmarking, testing and reporting to authorities assuming great importance in schools today. On the one hand, teachers have been given information from educators on what they should be doing regarding assessment. On the other hand, ideologically-driven bureaucrats have become more prescriptive by deciding what teachers will do. The reality for students is one of almost complete exclusion from the assessment process. The overwhelming view is that, in form and design, assessment tasks should not involve students: bureaucrats have a role, teachers have a scaled-down role, students have no role.

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 students and their parents what is really important for them 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).

## Design of present study

### Research objectives

The present research had three objectives:

- to establish clusters of students based on their perceptions of assessment,
- to describe the clusters obtained in the cluster solution, and
- to identify the key attributes of students who has negative perceptions of assessment tasks.

### Sample

A total of 3,055 student from primary and secondary schools in Queensland and Western Australia responded to a questionnaire seeking perceptions of impediments to leadership succession. Table 1 describes the sample which consisted of 2,038 primary school students and 1,017 secondary school students.

**Table 1**  
**Description of sample**

| Year    | Sample Size |        |                   |        | Total |
|---------|-------------|--------|-------------------|--------|-------|
|         | Queensland  |        | Western Australia |        |       |
|         | Male        | Female | Male              | Female |       |
| Year 6  | 230         | 287    | 334               | 200    | 1,051 |
| Year 7  | 158         | 196    | 345               | 288    | 987   |
| Year 8  | 88          | 95     | 148               | 152    | 483   |
| Year 9  | 59          | 72     | -                 | -      | 131   |
| Year 10 | 150         | 147    | 48                | 58     | 403   |
| Total   | 685         | 797    | 875               | 698    | 3,055 |

## Assessing Students’ Perceptions of Assessment

Students’ perceptions of assessment were assessed with the 30-item *Students’ Perceptions of Assessment Questionnaire (SPAQ)*. These items are assigned to five internally consistent scales. Table 2 shows these scales, their descriptions and sample items. The SPAQ is the result of instrument development and validation procedures conducted in Essex, England (DORMAN & KNIGHTLEY, 2006a; 2006b) and Australia (FISHER, WALDRIP, & DORMAN, 2005). Discussion of this earlier work is outside the scope of the present paper. The present form of the SPAQ employs a four-point Likert response format for each item (viz. Almost Never, Sometimes, Often, and Almost Always).

### Data analysis

Hierarchical cluster analysis – a procedure which attempts to identify relatively homogeneous groups of cases based on selected characteristics – was performed on the data. In the present study, this analysis was designed to establish clusters of respondents based on their perceptions of assessment tasks. To verify that the selected cluster solution separated the cluster groups, a multivariate analysis of variance (MANOVA) was performed on the data using the five SPAQ scales as dependent variables and cluster membership as the grouping variable.

To assist with the identification of key attributes of students who has negative perceptions of assessment tasks, a series of cross-tabulations involving the cluster variable and three categorical variables (viz. year, gender, class and state) were performed. Chi square tests to investigate departure of observed scores from expected scores were conducted for each cross-tabulation.

## RESULTS

### Validation of SPAQ

Reliability coefficients (Cronbach coefficient alpha) were computed for each SPAQ scale (see table 3). These results show that all scales had at least satisfactory internal consistency. Indices ranged from .63 for Diversity to .83 for Authenticity. Table 3 also shows means, standard deviations and indices for skewness and kurtosis. Some departure from normality was evident with 4 of the 5 scales having statistically significant skewness and kurtosis ( $p < .05$ ).

**Table 2**  
**Descriptive information for five SPAQ scales**

| Scale                            | Scale description   | Sample item  |
|----------------------------------|---|--|
| Congruence with Planned Learning | The extent to which assessment tasks align with the goals, objectives and activities of the learning program. | My assignments/tests are about what I have done in class.                    |
| Authenticity                     | The extent to which assessment tasks feature real life situations that are relevant to the learner.           | I find science assessment tasks are relevant to what I do outside of school. |
| Student Consultation             | The extent to which students are consulted and informed about the forms of assessment tasks being employed.   | I have a say in how I will be assessed in science                            |
| Transparency                     | The extent to which the purposes and forms of assessment tasks are well-defined and clear to the learner.     | I am clear about what my teacher wants in my assessment tasks.               |
| Diversity                        | The extent to which all students have an equal chance at completing assessment tasks.                         | I have as much chance as any other student at completing assessment tasks.   |

**Table 3**  
Validation data and scale statistics for five SPAQ scales  
(N = 3,055 students in 149 classes)

| Scale                               | Coefficient<br>$\alpha$ | Mean<br>Correlation | Mean  | Standard<br>Deviation | Skewness | Kurtosis |
|-------------------------------------|-------------------------|---------------------|-------|-----------------------|----------|----------|
| Congruence with<br>Planned Learning | .73                     | .36                 | 18.40 | 3.22                  | -.41*    | -.13     |
| Authenticity                        | .83                     | .42                 | 14.11 | 3.95                  | .12*     | -.48*    |
| Student Consultation                | .72                     | .44                 | 13.47 | 3.64                  | .37*     | -.18*    |
| Transparency                        | .82                     | .45                 | 18.23 | 3.88                  | -.51*    | -.28*    |
| Diversity                           | .63                     | .44                 | 15.40 | 3.38                  | -.02     | -.22*    |

\*  $p < .05$

Discriminant validity for each SPAQ scale was explored through the mean correlation of the scale with the remaining four scales. The results shown in table 3 indicate some scale overlap but not to the extent that would confound interpretation of results. Additionally, all scales should be retained because of their conceptual distinctiveness.

#### Cluster Analysis

A review of dendrograms based on hierarchical cluster analysis indicated that a four cluster solution with 3,027 students from the sample would be appropriate. These four homogeneous groups (Clusters 1, 2, 3 and 4) contained 745, 831, 645 and 806 respondents respectively. Table 4 shows descriptive statistics for these four Clusters. Consideration of the mean scores for these clusters indicated the following: scale means for Clusters 1 and 2 were mid-range – between the means for Clusters 3 and 4. However, Clusters 1 and 2 can be distinguished by the means for two scales: Authenticity and Transparency. Whereas Cluster 1 had medium Authenticity and Transparency, Cluster 2 had low Authenticity and high Transparency. Cluster 3 had low mean scores for all SPAQ scales. That is, Cluster 3 respondents did not perceive assessment tasks positively. By contrast, Cluster 4 had high mean scores for all SPAQ scales.

**Table 4**  
Means and Standard Deviations for the four cluster solution  
(N = 3,027 students)

| SPAQ Scale                          | Cluster        |      |                |      |                |      |                |      |
|-------------------------------------|----------------|------|----------------|------|----------------|------|----------------|------|
|                                     | 1<br>(n = 745) |      | 2<br>(n = 831) |      | 3<br>(n = 645) |      | 4<br>(n = 806) |      |
|                                     | M              | SD   | M              | SD   | M              | SD   | M              | SD   |
| Congruence with<br>Planned Learning | 17.19          | 2.66 | 19.62          | 2.52 | 15.65          | 3.10 | 20.47          | 2.24 |
| Authenticity                        | 15.22          | 2.32 | 12.10          | 2.61 | 10.25          | 2.54 | 18.24          | 2.71 |
| Student Consultation                | 14.11          | 2.54 | 12.13          | 2.32 | 9.88           | 2.19 | 17.15          | 2.95 |
| Transparency                        | 16.40          | 2.31 | 20.41          | 2.09 | 13.52          | 2.79 | 21.46          | 2.05 |
| Diversity                           | 15.19          | 2.59 | 15.51          | 2.43 | 11.86          | 2.50 | 18.29          | 2.66 |

To verify this four cluster solution, a multivariate analysis of variance (MANOVA) was performed on the data using the five SPAQ scales as dependent variables and cluster membership as the grouping variable. This MANOVA was significant with Wilks' lambda criterion of 0.104 [ $F(15,$

8334) = 706.34 ( $p < .001$ )]. Univariate  $F$  tests for the effect of cluster grouping on each SPAQ scale yielded the following results: Congruence with Planned Learning,  $F(3, 3023) = 514.928$  ( $p < .001$ ); Authenticity,  $F(3, 3023) = 1,414.12$  ( $p < .001$ ); Student Consultation,  $F(3, 3023) = 1,088.12$  ( $p < .001$ ); Transparency,  $F(3, 3023) = 1,824.11$  ( $p < .001$ ); and Diversity,  $F(3, 3023) = 772.06$  ( $p < .001$ ). These analyses confirm that this cluster solution separated the respondents into four distinct groups. Tukey's post-hoc procedure revealed that all 24 pairwise cluster comparisons were significant ( $p < .05$ ). Effect sizes were computed using Cohen's (1977)  $d$  (the difference between group means per full sample standard deviation) as a convenient index. These values ranged from 0.95 for Diversity (Clusters 1 and 2) to 2.06 for Transparency (Clusters 3 and 4) and can be taken as large to very large in the context of social science research. It is noteworthy that effect sizes for all five scales for comparisons of Clusters 3 and 4 were very large: Congruence with Planned Learning,  $d = 1.50$ ; Authenticity,  $d = 2.03$ ; Student Consultation,  $d = 2.00$ ; Transparency,  $d = 2.06$ ; and Diversity,  $d = 1.91$ .

The third research objective focussed on the characteristics of Cluster 3 – those respondents who perceived negative perceptions of assessment tasks. To facilitate discussion on this issue, separate cross-tabulations involving cluster membership with year group and gender were conducted. Table 5 shows, for each grouping variable, the percentage of respondents that fell in each of the four clusters. For example, 25.6% of Year 6 students and 19.1% of Year 7 students fell in Cluster 1. Chi square tests conducted on each of these cross-tabulations indicated significant differences between the observed results and expected results for year group,  $\chi^2(3, N = 3,027) = 238.44$  ( $p < .001$ ) and state,  $\chi^2(3, N = 3,027) = 134.47$  ( $p < .001$ ), but not gender,  $\chi^2(3, N = 3,027) = 6.27$  ( $p = .10$ ).

Interpretation of the data in table 5 suggests the following conclusions. The distribution of students in each cluster varies according to year group. For example, 19.1% of Year 7 students and 41.5% of Year 9 students were in Cluster 1. Cluster 3 members held negative perceptions of assessment and its membership ranged from 13.8% of Year 9 students to 27.9% of Year 8 students. Conversely, only 13.4% of Year 8 students but 38.5% of Year 9 students were located in Cluster 4 (positive perceptions of assessment). The distribution of students across the clusters is relatively similar for male and female students. For example, 21.5% of males and 21.1% of females were assigned to Cluster 3. Finally, a significantly higher proportion of Queensland students were assigned to Cluster 4 (34.2%) compared to Western Australian students (19.7%).

To establish a general profile of students who had negative perceptions of assessment it was necessary to focus on the members of Cluster 3. A description of this cluster and the full sample used in this study in terms of year level, gender, and state is shown in Table 6. For example, out of a total sample of 230, 39 male Year 7 students from Queensland were assigned to Cluster 3.

**Table 6**  
Description of Cluster 3 and full sample

| Year  | State      |           |                   |           |
|-------|------------|-----------|-------------------|-----------|
|       | Queensland |           | Western Australia |           |
|       | Male       | Female    | Male              | Female    |
| 6     | 39 (230)   | 31 (287)  | 56 (334)          | 39 (200)  |
| 7     | 29 (158)   | 50 (196)  | 99 (345)          | 88 (288)  |
| 8     | 22 (88)    | 29 (95)   | 46 (148)          | 36 (152)  |
| 9     | 10 (59)    | 8 (72)    | 0 (0)             | 0 (0)     |
| 10    | 20 (150)   | 15 (147)  | 13 (48)           | 13 (58)   |
| Total | 120 (685)  | 133 (797) | 214 (875)         | 176 (698) |

**Table 5**  
Percentage of year group, gender and state for four clusters  
(N = 3,027 students)

| Cluster | Year Group |        |        |        |         | Gender |        | State |      |
|---------|------------|--------|--------|--------|---------|--------|--------|-------|------|
|         | Year 6     | Year 7 | Year 8 | Year 9 | Year 10 | Male   | Female | Qld   | WA   |
| 1       | 25.6       | 19.1   | 20.8   | 41.5   | 35.4    | 26.0   | 23.1   | 27.6  | 21.8 |
| 2       | 25.8       | 32.6   | 37.9   | 6.2    | 15.5    | 25.7   | 29.4   | 20.7  | 33.7 |
| 3       | 16.1       | 27.4   | 27.9   | 13.8   | 15.8    | 21.5   | 21.1   | 17.5  | 24.8 |
| 4       | 32.5       | 20.9   | 13.4   | 38.5   | 33.3    | 26.8   | 26.4   | 34.2  | 19.7 |

These data indicate that all year 7 and 8 student groups apart from male students in Queensland had proportions in Cluster 3 above 0.20. The highest proportion was for male year 8 students in Western Australia (46/148). Additionally, year 10 students in Western Australia featured highly in Cluster 3 (13/48 for males and 13/58 for females). The lowest group membership for Cluster 3 was for female Year 10 students in Queensland (15/147).

## DISCUSSION

The research reported in this paper is important to school assessment for at least four reasons. First, the study demonstrates the usefulness of the SPAQ to assess students' perceptions of assessment. It builds upon and extends earlier work by Dorman and Knightley (2006a). While this study shows the SPAQ's sound structural characteristics, it should be cross-validated with other samples and in other countries. Such work would enhance the utility of the SPAQ. One particularly useful direction would be to use translations of the SPAQ in countries where English is not the first language.

Second, the study has revealed substantial variation in how students perceive assessment tasks. While 26.62% of the assigned sample of 3,027 students were in Cluster 4 (positive perceptions of assessment), 21.31% were in Cluster 3 (negative perceptions of assessment). Teachers need to be aware that students do not view assessment tasks in a uniform manner. The next step in this research agenda should be to ascertain whether particular tasks are associated with positive and negative perceptions of assessment.

Third, different demographic groupings were disproportionately represented in some clusters. In particular, Cluster 3 (negative perceptions of assessment) had a high proportion of year 10 and male year 8 students in Western Australia. Further research employing case study techniques is needed to identify the reasons for these results. Can these findings be related to specific assessment activities being undertaken by students in Western Australia?

Finally, this research raises questions about the professional development activities relating to assessment that are available to school staff. According to the Assessment Reform Group (2000), assessment for learning should be a key professional skill of teachers. Popham (2006) contrasts assessment *for* learning with assessment *of* learning. Assessment of learning attempts to identify what students know for the purposes of giving grades or evaluating schools. Assessment for learning is always about *what's next instructionally*? Policy developers need to recognise this very clear distinction and ensure that assessment is not simply for people outside the classroom. Indeed, high stakes testing geared at meeting bureaucratic and political needs does not improve student motivation and can increase boredom, fear, hostility and disillusionment in students (see Amrein & Berliner, 2003; Sacks, 1999).

Within this discussion it is particularly important to recognise that attempts to employ accountability measures that decontextualise student learning reflect managerial rather than educational imperatives. LINN (2000) makes this point clear when asserting that assessment-based accountability models have not been shown to improve education. While the present paper focuses on characteristics of assessment tasks within teaching and learning cycles (Assessment Reform Group, 2002; HATTIE & JAEGER, 1998), bureaucratic accountability measures deflect attention from students and classrooms. Indeed, as MAYER (2005) has recently asserted, bureaucratic models of teacher accountability are not improving student learning, especially for those most at risk. The research presented in this paper intentionally focused on issues that classroom teachers can address. It has the potential to provide a sound foundation for authentic assessment in schools.

## CONCLUSIONS

This paper has reported the use of a relatively new instrument, the Students' Perceptions of Assessment Questionnaire (SPAQ) in middle school science classes. Despite the reality that students have the best vantage point for assessing assessment tasks, much deliberation on assessment is undertaken away from the classroom and by people who believe that uniform sets of behaviours of teachers and students are essential to improved outcomes. By eliciting students' high inference summary judgments rather than external observers/ evaluators, this research parallels the methodology in science classroom environment research pioneered by Walberg (1976) in the 1960s. Student perceptual data should be used to assess the quality of assessment tasks that students perform and it is hoped that the SPAQ will be used in cross-national research in this area.

## BIBLIOGRAPHY

Amrein, A.L., & Berliner, D.C. The effects of high-stakes testing on student motivation and learning, *Educational Leadership* 60, (5), 32-38, 2003.

- Assessment Reform Group, *Assessment for learning: Beyond the black box*. Retrieved March 19, 2007, from <http://arg.educ.cam.ac.uk/AssessInsides.pdf>, 1999.
- Assessment Reform Group, *Assessment for learning: 10 principles*. Retrieved March 19, 2007, from <http://arg.educ.cam.ac.uk/CIE3.pdf>, 2002.
- Barksdale-Ladd, M.A. & Thomas, K.F. What's at stake in high-stakes testing: teachers and parents speak out, *Journal of Teacher Education*, 51, (5), 384-397, 2000.
- Black, P. & Wiliam, D. Assessment and classroom learning, *Assessment in Education*, 5, 7-74, 1998.
- Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (Rev. ed.), New York, Academic, 1977.
- Crooks, T.J. The impact of classroom evaluation practices on students, *Review of Educational Research*, 58, (4), 438-481, 1988.
- Dorman, J.P. & Knightley, W.M. Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks, *Educational Studies*, 32, (1), 47-58, 2006a.
- Dorman, J.P. & Knightley, W.M. Initial use of the Perceptions of Assessment Tasks Inventory (PATI) in English secondary schools, *Alberta Journal of Educational Research*, 52, (3), 196-199, 2006b.
- Fisher, D.L.; Waldrup, B.G. & Dorman, J.P. *Student perceptions of assessment: Development and validation of a questionnaire*. Paper presented at the annual meeting of the American Educational Research Association (AERA), Montreal, 2005, April.
- Harlen, W. Teaching for understanding in pre-secondary science, In B.J. Fraser & K. G. Tobin (eds.), *International Handbook of Science Education* (pp. 183-198). Dordrecht, The Netherlands: Kluwer, 1998.
- Hattie, J. & Jaeger, R. Assessment and classroom learning: A deductive approach, *Assessment in Education*, 5, (1), 111-122, 1998.
- Linn, R.L. Assessments and accountability, *Educational Researcher*, 29, (2), 4-16, 2000.
- Mayer, D. Reviving the "policy bargain" discussion: Professional accountability and the contribution of teacher - performance assessment, *The Clearing House*, 78, (4), 177-181, 2005.
- Plake, B.S. Teacher assessment literacy: Teachers' competencies in the educational assessment of students, *Mid-Western Educational Researcher*, 6, (1), 21-27, 1993.
- Popham, W.J. Consequential validity: Right concern-wrong concept. *Educational Measurement: Issues and Practice*, 16, (2), 9-13, 1997.
- Popham, W.J. Assessment for learning: An endangered species, *Educational Leadership*, 63, (5), 82-83, 2006.
- Reynolds, D.S.; Doran, R.L.; Allers, R.H. & Agruso, S.A. *Alternative assessment in science: a teacher's guide*, Buffalo, NY, University of Buffalo, 1995.
- Rogoff, B. Student assessment for the information age, *The Chronicle of Higher Education*, 48, (3), B17, 2001, September 14.
- Sachs, J. Reclaiming the agenda of teacher professionalism: An Australian experience, *Journal of Education for Teaching*, 23, (3), 263-276, 1997.
- Sacks, P. *Standardised Minds: The High Price of America's Testing Culture and What we Can do to Change It*, Cambridge, MA, Perseus, 1999.
- Schaffner, M.; Burry-Stock, J.A.; Cho, G.; Boney, T. & Hamilton, G. *What do kids think when their teachers grade?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, 2000, April.
- Steinberg, J. Student failure causes states to retool testing programs. *The New York Times*, p. A1, 2000, December 22.
- Stiggins, R., *Student-centered Classroom Assessment*. Ontario: Macmillan College Publishing, 1994.
- Walberg, H.J. Psychology of learning environments: Behavioral, Structural, or perceptual? *Review of Research in Education*, 4, (1), 142-178, 1976.

## APPENDIX

### Students' Perceptions of Assessment Questionnaire (SPAQ)

- Questions in science tests what I know.
- My science assignments/tests examines what I do in class.
- My assignments/tests are about what I have done in class.
- How I am assessed is like what I do in class.
- How I am assessed is similar to what I do in class.
- I am assessed on what the teacher has taught me.
- I am asked to apply my learning to real life situations.
- My science assessment tasks are useful in everyday things.
- I find science assessment tasks are relevant to what I do outside of school.
- Assessment in science tests my ability to apply what I know to real-life problems.
- Assessment in science examines my ability to answer every day questions
- I can show others that my learning has helped me do things.
- 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

**Received: 20.03.2007**  
**Approved: 29.09.2007**