*The Alberta Journal of Educational Research* *Vol. 55, No. 4, Winter 2009, 435-452*

*Linda M. Phillips*
*Stephen P. Norris*
*Martha L. Smith*
*Jodi Buker*
and
*Chandra Kasper*
*University of Alberta*

# Assessment Techniques Corresponding to Scientific Texts in Commercial Reading Programs: Do They Promote Scientific Literacy?

*This research is part of a larger study of commercial reading programs used in Canada in grades 1-6. The specific purposes of the results reported here were to identify and quantify the assessment techniques suggested for the selections that contain scientific content, to show how the assessments differ by grade, to evaluate the nature and quality of the assessments, and to examine the extent to which the assessments help foster scientific literacy. It was found that the assessments occurred in six major forms and employed about a dozen assessment tools that engage students in nearly 20 tasks. Such variety is endorsed in both literacy and science education position statements. The assessments showed some weak trends by grade, but primarily left the purpose of the assessments to teachers' judgment. The consequence is that teachers probably will choose the assessments for formative rather than summative evaluation, an approach also endorsed by literacy and science education policy statements. Hardly any of the assessments focused on the specificities of learning to read texts that are scientific such as interpreting descriptions of methods and research findings and thus had limited use in promoting this particular aspect of scientific literacy.*

*Cette recherche s'inscrit dans le cadre d'une plus grande étude portant sur les programmes de lecture commerciaux qu'on emploie au Canada de la 1re à la 6e année. Les objectifs précis de l'étude étaient d'identifier et de quantifier les techniques d'évaluation proposées pour les sélections avec un contenu scientifique; de démontrer les différences entre les évaluations d'une année à l'autre; d'évaluer la nature et la qualité des évaluations; et d'examiner la mesure dans laquelle les évaluations contribuent à la littératie scientifique. Les résultats indiquent qu'il existe six formes de l'évaluation et qu'elle implique environ une douzaine d'outils d'évaluation qui font participer les étudiants à presque 20 tâches. Les milieux*

---

Linda Phillips is a professor of language and literacy studies and Director of the Canadian Centre for Research on Literacy, Department of Elementary Education.
Stephen Norris is a professor and Canada Research Chair in Scientific Literacy and the Public Understanding of Science, Department of Educational Policy Studies.
Martha Smith is a contract researcher and a freelance writer/editor. She was a research associate in the Canadian Centre for Research on Literacy while this study was being conducted.
Jodi Buker is a program coordinator for the Read With Me program in the Calgary Public Library. At the time of this study, she worked in the Canadian Centre for Research on Literacy for the duration of her Roger Smith Undergraduate Research Award.
Chandra Kasper is a teacher with Edmonton Public Schools. She was working as a research assistant in the Canadian Centre for Research on Literacy while this study was being conducted.

*éducatifs, notamment les domaines de la littératie et des sciences, appuient une telle variété. Les évaluations ont révélé des tendances indiquant certaines faiblesses par niveau scolaire, mais ont surtout laissé aux enseignants la tâche de juger du but des évaluations. Par conséquent, les enseignants choisiront probablement les évaluations pour des raisons formatives plutôt que sommatives, ce qui constituent également une approche qu'appuie l'enseignement de la littératie et des sciences. Presqu'aucune des évaluations ne portaient sur les spécificités de l'apprentissage de la lecture de textes scientifiques, telles que l'interprétation des descriptions de méthodologie et des résultats de recherche, limitant ainsi leur utilité dans la promotion de cet aspect précis de la littératie scientifique.*

We present an analysis of the assessment techniques corresponding to scientific texts contained in three of Canada's most widely used commercial reading programs for grades 1-6. Although previous generations of basals were dominated by narrative content (Flood & Lapp, 1987; Moss & Newton, 2002; Murphy, 1991), publishers of today's reading programs include more text with informational qualities (Phillips, Smith, & Norris, 2005), a significant portion of which is science content (Smith, Phillips, Norris, Guilbert, & Stange, 2006). The inclusion of this material has potential relevance for developing scientific literacy in school children (Phillips & Norris, 2003; Smith et al., 2006). In addition, an orientation toward hands-on learning during science lessons themselves can mean that little instruction in the reading of scientific text occurs in elementary science lessons (Heselden & Staples, 2002) and that the responsibility for such instruction might be seen to fall to literacy lessons that rely heavily on the use of commercial reading programs.

Assessment plays a crucial role in both literacy learning (Johnston & Costello, 2005) and science education (National Science Teachers Association [NSTA], 2001), yet the assessment techniques suggested in the selections of commercial reading programs that contain scientific content have not been documented, analyzed, or evaluated. Therefore, in the study reported here, we focus on the nature of the assessments accompanying scientific material in our examination of the role of these programs in fostering scientific literacy in elementary schoolchildren.

## *Background*

The science-literacy connection is increasingly being explored as a way of developing both literacy and science content knowledge. The term *scientific literacy* is broad, and although many shades of meaning coexist, it is generally used to mean being knowledgeable about or learned and educated in science (Norris & Phillips, 2003). A growing number of researchers have, however, argued for the importance of language and the role of literacy instruction in science education (Palinscar & Magnusson, 2001; Saul, 2004; Wellington & Osborne, 2001). We have advanced the view that learning through text has a significant role in developing scientific literacy (Norris & Phillips; Norris et al., 2008), and this perspective is foundational to the investigation reported here.

The rationale for this study builds on findings from both science education and the literacy field. Learning from text has not been a focus of current primary and elementary science education for a variety of reasons. One explanation lies in teachers' perceptions of science education and their knowledge of science (Brickhouse, 1990; Pajares, 1992; Pomeroy, 1993). Some research has suggested that elementary teachers have weak understandings of science (Harlen, 1997; Newton & Newton, 2000; Trumper, 2003), and many do not see reading science as different from reading any other subject (Pappas, 2006; Shymansky, Yore, & Good, 1991). In addition, teachers have widely embraced a constructivist philosophy that emphasizes first-hand, student-centered learning; hands-on activities; and the fostering of positive attitudes toward science (Levitt, 2001). Although this orientation has obvious benefits, it can also lead to a diminished expectation that print is an important source of science for students (Rowell & Ebbers, 2004). Official documents may reinforce this perspective: according to Vanderwolf, Cook, Coutts, and Cropp (2005), Canadian science curricula (K-12) emphasize discovery learning and group work while downgrading the use of textbooks. Thus it would not be surprising to find, as Palinscar and Magnusson (2001) did, that text is often neglected when activity-based inquiry practices are embraced by teachers in science education and that attention to text has come mainly from the reading education community.

In the literacy field, ongoing efforts promote the ideal of content-area teachers being teachers of reading, but the evidence suggests a resistance to literacy instruction in the science domain. Thus fostering the fundamental sense of scientific literacy (Norris & Phillips, 2003), by which we mean the ability to read and write when the content is science, generally falls to teachers who provide reading instruction. Many teachers rely on commercial reading programs for much of their reading instruction (Morrow & Gambrell, 2000; Smith, Phillips, Leithead, & Norris, 2004). These programs have been shown to have a literary emphasis (Flood & Lapp, 1987; Moss & Newton, 2002; Murphy, 1991), but recent research (Phillips, Smith, & Norris, 2005) indicates some change in this regard toward more and earlier exposure to informational text. Canada's current programs contain a variety of text types, including about a quarter with scientific content. Because the inclusion of this material conveys an impression of usefulness for supporting the teaching of scientific literacy, it is important to investigate the extent to which this is the case. In an earlier investigation (Smith et al., 2006), we looked at the quantity and nature of scientific texts included in three widely used commercial reading programs and examined the nature and quality of reading instruction corresponding to these texts. This study builds on that investigation, but focuses specifically on assessment.

Assessment is intimately related to instruction and learning. In the literacy field, there is widespread agreement that students experience more difficulty with expository texts (e.g., content area material) than with literary texts (Alvermann & Boothby, 1982; Graesser, McNamara, & Louwerse, 2003; Saenz & Fuchs, 2002), and in the science education field it is "no secret that students find their science texts difficult to read" (Shanahan, 2004, p. 370). Research shows that without effective reading instruction, competence in understanding and critically analyzing scientific expository text is seriously compromised (Craig & Yore, 1996; Norris & Phillips, 1994; Otero & Companario, 1990; Phillips & Norris, 1999). Assessment plays an important role in the effectiveness of such instruction. The *what* and *how* of assessment have implications for what is taught and learned (Johnston & Costello, 2005), and consequently, the importance of assessment for learning is well recognized in both the fields of science education (NSTA, 2001) and literacy (IRA & NCTE, 1994). Although commercial reading programs may be expected to include general literacy assessments, such assessment must also implicate science content and reasoning if it is to contribute at all significantly to the development of scientific literacy. As Pappas (2006) has argued, science is a particular discipline with distinctive uses of language. Thus assessments must reflect this distinctiveness by, for example, probing for students' knowledge of and facility with the technical language and vocabulary of science. Much more is involved, however, because assessments must also examine students' ability to interpret the reported purposes of scientific work and the motivations of scientists in carrying it out; the described methods of science and how they relate to purposes; the conclusions of scientific work; and perhaps most important of all, the reasoning that holds together purposes, methods, and conclusions. If the focus of these earlier assessments can be said to be on the macroscopic aspects of scientific texts, there is also the need for assessments that focus on microscopic aspects. Such microscopically focused assessments must attend to students' abilities to determine the reputed truth status of scientific statements (e.g., true, likely true, uncertain, doubtful, false); the role of statements in scientific reasoning (e.g., descriptions of method, reports of results, descriptions of conclusion); and the status of statements (e.g., observations, causal relations, correlational statements). All such assessments are of comprehension and interpretation, but specifically tuned to the scientific context. To evaluate the usefulness of the programs' assessments in this regard, this study had the following purposes: (a) identification and quantification of assessments in the teaching units that contain science content, (b) examination of whether or how the suggested assessment techniques differ by grade, (c) examination of the nature or quality of the assessments, and (d) an evaluation of the importance of these assessments for fostering scientific literacy.

## *Methodology*

### *Data Sources*

The currently and most widely used commercial reading programs in Canada were identified through communication with the Ministries of Education for all provinces and territories. Each jurisdiction identified at least one of the following, which we refer to as Gage, Ginn, and Nelson: (a) *Cornerstones Canadian Language Arts* by Gage (1998-2001), (b) *Collections* by Prentice Hall Ginn Canada (1996-2000), and (c) *Nelson Language Arts* by Nelson Thomson Learning (1998-2001). We obtained complete program sets for grades K-6 from the respective publishers.

Program sets contained student books (anthologies), teachers' guides, and various ancillary materials for each grade. Across the three publishers, the six grades included 72 student books that we inventoried by selections contained therein. The instructional units in the accompanying teachers' guides also were inventoried. In total there were 1,106 selections intended for use in 980 instructional units.

The subset of selections with science content was identified earlier (Smith et al., 2006). These 238 selections were contained in 233 instructional units, which became the units of analysis for this study. Although publishers did have general assessment handbooks, this study focused on the actual suggestions in the instructional units as these are tailored to the selections and are the assessments most likely to be used with these selections.

*Procedure*

Neuendorf (2002) was used as a guide for content analysis of the identified instructional units. We began determining the diversity of assessment suggestions in the programs by sampling units at all grades from the entire inventory, progressively developing lists of assessment types encountered. As identified instances began to fit previously identified types, we devised a preliminary code form and code book containing explanations and exemplars. Regular meetings were held to compare notes, raise issues, and refine the protocol. The final measurement protocol was developed through an iterative process involving several rounds of pilot reliability testing as well as examination and discussion of disagreements leading to modification of the coding scheme. When high levels of inter-rater reliability had been achieved and all team members expressed agreement on the protocol, two trained coders were assigned to alternate cases (i.e., one coder was responsible for odd-numbered cases, the other for even-numbered cases). The 233 teaching units were thus divided, and approximately the same number of units for each grade and each publisher fell to the coders.

*Identification of Assessment Instances*

Instances of assessment were identified in three ways. Some were clearly labeled in various side notes accompanying suggested student tasks in the teachers' guides. For example, the contents of side boxes or sections labeled "Assess Learning" or "Ongoing Assessment" were considered to be clear instances of assessment. Other instances were not labeled, but from explicit directions to the teacher, it could be inferred that assessment was intended. Examples of this include the use of words like *sample criteria, tests/quizzes, evaluation,* or *feedback*. Directions such as "Record these as criteria that students can use to guide their work and offer constructive peer feedback" (Nelson 4, *Times to Share*, p. 130) imply assessment as do references to pre- and post-spelling tests. A third method of identification involved locating implicit mentions of assessment. For example, the directions "Ask the other group members to listen closely to determine such things as whether: the question was addressed; the answer was correct; sufficient details, explanations or proofs were provided; the answer was easy to understand" (Ginn 6, *Space, Stars, and Quasars*, p. 34) indicated an implicit instance of peer assessment. Instructional units could contain multiple instances of assessment. Instances were coded sequentially in the order in which they were encountered in each unit to facilitate reliability analyses.

*Major Dimensions of Investigation:*

Each instance of assessment was coded along six major dimensions: *Form, Tool, Assessment Task, Purpose, Correspondence to Learning Objectives,* and *Content Area*. For each dimension we developed coding categories supplemented with explanations and exemplars. These dimensions served as areas for analysis for summary data (across all grades and publishers) and grade data.

Main assessment *forms* were identified as Observation, Student Self-Assessment, Peer-Assessment, Conference, and Product/Skill Evaluation. A category of Unknown was used for coding all other instances that did not fit into one of the above categories or for which the main form of assessment was unclear. Suggestions for portfolio inclusion were also collected and coded, but these were not considered instances of assessment in this analysis.

The *tool* was the recording device used to keep track of the information gained from the assessment and was coded as one of the following: Checklist, Rating Scale, Anecdotal Notes, Running Record, Answer Key/Marking Guide, Journal/Learning Log Entry, Criteria, Guiding Questions/Prompts, Other, Combinations of any of these, and Not Specified (which included cases where an informal method of recording, such as in the teacher's head, is assumed).

*Task* denotes the activity or product that corresponds to the assessment instance. In some cases, the activity description calls for the student(s) to do more than one task. These were coded as Combinations of Tasks although the assessment may not correspond to all tasks. Eighteen task categories were coded, as outlined in the *Results* section.

*Purpose* of the assessment refers to how the program suggested that the assessment should be used to guide instruction (at a general level): Summative, Formative, or Diagnostic. A combination category was used for cases in which the directions for the teacher suggested several simultaneous purposes. Cannot Discern was a classification given when the program did not give an indication to the teacher of how to use the information gained from the assessment.

For each instance of assessment encountered, the relatedness to stated learning objectives was coded. This *correspondence* was categorized in one of three ways: Corresponds, Does not Correspond, Unrelated. The first two codes indicated whether the assessment was judged to be a good indicator of student achievement of given learning objectives, and the third was reserved for cases in which there was no clear learning objective corresponding to the assessment.

*Content area* was coded for each assessment instance. We assumed that each instance contained literacy content involving assessment of the knowledge or skills required to use one or more of the six strands of language arts outlined in all Canadian curricula: reading, writing, listening, speaking, viewing, and representing. We distinguished, however, assessments implicating language arts content only versus those that also implicated other subject areas including science, social studies, mathematics, technology, music, drama, art, and health. In addition, coding included Cross-curricular Skills, Interpersonal Skills, and Combinations (including science and other). A category of Cannot Discern was used when the coder could not determine a particular content area that was implicated in the assessment either when considered on its own or in the context of the student task that was being assessed. We note that although all the student selections contained scientific content, because this was the basis for choosing them, hardly any were totally science. About two thirds of the selections were judged to be between 88% and 100% scientific content (Norris et al., 2008), with the remaining selections containing a smaller proportion of science. Therefore, it was frequently the case that assessment guidance referred to content areas other than science. We report in the *Results* on all the content areas covered. The significance of reporting in this way is to indicate the degree to which science content and reasoning is the focus of assessment in selections that are known to contain science.

*Reliability of Assessment Coding*
A 10% random sample stratified by grade and publisher was selected from the 233 coded units. Each chosen unit was assigned to the coder who had not already coded the case. The two coding forms for each of the reliability cases were compared. Agreements and disagreements on the codes for each assessment instance were tallied, compiled across cases, and used to calculate total percentage agreements. There was 92% agreement on the number of assessments in the units, 97% agreement on the identification of assessment instances, and a range of 74% (content area) to 97% (form) agreement on the six major dimensions of the investigation. The low percentage agreement on content area reflects the difficulty in determining relevance to specific subject areas: most (84%) disagreements involved a judgment of either Cannot Discern or Cross-curricular Skills, although the other coder made a more specific determination (e.g., language arts only, interpersonal skills, etc.). The average percent agreement on the six dimensions was 85%, and the overall average for all variables used in the analyses was 87%.

## *Results*

*Instances of Assessment*
A total of 704 instances of assessment were identified. Although the number of instances per unit ranged from 0-9, most units (92%) contained at least one assessment. On average, there were three instances per unit: over half of the units (55%) had 1-3 assessments, and only 13% had more than five. Over half (54%) of the assessments were identified through the presence of labels such as Assess Learning. The remainder were identified through explicit direction in 34% of cases and implicitly through wording in 12% of cases.

*Nature of the Assessments*
The nature of assessments was examined through analysis of the six major dimensions outlined above. We analyzed the overall data (across grades and publishers) for each dimension first and then looked for any differences by grade.

Table 1 outlines the results on *form*. Overall, observation was the most frequent main form of assessment, accounting for over a third (38%) of all identified instances. Some similarities and differences from the overall pattern were found by grade. At all grades, observation and student self-assessment were the most frequently occurring forms of assessment, together making up 54%-84% of all assessments in each of the six grades. Grades 1 and 2, however, had higher rates of observation (64% and 52% respectively). The frequency of this form of assessment then drops to approximately 30% of assessment at each grade. Grade 3 appears to be pivotal in several other respects as well. Although far less frequent than observations at any grade, conferences are most frequent in the first two grades. The decline in observations and conferences at grade 3 and beyond was accompanied by an increase in several other forms of assessment. Self-assessment, which is consistently high, showed an increase at grade 3, as did the initially infrequent forms of peer assessments and product/skill evaluations.

Table 1: Frequencies and Percentages of Forms of Assessment by Grade

| | Grade 1 | | Grade 2 | | Grade 3 | | Grade 4 | | Grade 5 | | Grade 6 | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Form of Assessment* | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % |
| Observations | 70 | 64 | 59 | 52 | 37 | 29 | 32 | 26 | 36 | 31 | 37 | 32 | 271 | 38 |
| Student Self-Assessments | 22 | 20 | 18 | 16 | 38 | 30 | 34 | 28 | 31 | 27 | 25 | 22 | 168 | 24 |
| Product/Skill Evaluation | 2 | 2 | 9 | 8 | 21 | 17 | 25 | 20 | 20 | 17 | 15 | 13 | 92 | 13 |
| Peer-Assessments | 5 | 5 | 7 | 6 | 18 | 14 | 14 | 11 | 15 | 13 | 18 | 16 | 77 | 11 |
| Unknown/Other | 3 | 3 | 10 | 9 | 7 | 6 | 12 | 10 | 10 | 9 | 15 | 13 | 57 | 8 |
| Conference | 7 | 6 | 11 | 10 | 6 | 5 | 5 | 4 | 4 | 3 | 6 | 5 | 39 | 6 |
| Total per Grade | 109 | 100 | 114 | 101 | 127 | 101 | 122 | 99 | 116 | 100 | 116 | 101 | 704 | 100 |

Table 2: Frequencies and Percentages of Tools Used in Assessment by Grade

| | Grade 1 | | Grade 2 | | Grade 3 | | Grade 4 | | Grade 5 | | Grade 6 | | Total | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Assessment Tools* | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % | *f* | % |
| Guiding Questions | 44 | 40 | 53 | 46 | 61 | 48 | 51 | 42 | 52 | 45 | 47 | 41 | 308 | 44 |
| Answer Key | 2 | 2 | 8 | 7 | 26 | 20 | 28 | 23 | 18 | 16 | 18 | 16 | 100 | 14 |
| Unspecified | 21 | 19 | 19 | 17 | 10 | 8 | 18 | 15 | 9 | 8 | 18 | 16 | 95 | 13 |
| Criteria | 1 | 1 | 11 | 10 | 11 | 9 | 15 | 12 | 18 | 16 | 19 | 16 | 75 | 11 |
| Combination | 14 | 13 | 13 | 11 | 12 | 9 | 5 | 4 | 7 | 6 | 7 | 6 | 58 | 8 |
| Checklist | 9 | 8 | 4 | 4 | 4 | 3 | 1 | 1 | 6 | 5 | 3 | 3 | 27 | 4 |
| Other | 7 | 6 | 2 | 2 | 1 | 1 | 0 | 0 | 4 | 3 | 1 | 1 | 15 | 2 |
| Anecdotal Notes | 2 | 2 | 0 | 0 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 9 | 1 |
| Running Record | 5 | 5 | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 1 |
| Rating Scale | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | 4 | 1 |
| Journal Entry | 4 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 1 |
| Total Per Grade | 109 | 100 | 114 | 101 | 127 | 100 | 122 | 101 | 116 | 101 | 116 | 102 | 704 | 100 |

Table 2 shows the occurrence of *tools* used in the assessment instances. Guiding questions were by far the most frequently used tools both overall (44%) and by grade (range 40-48%). Overall data showed that use of other tools is relatively infrequent. The use of criteria was relatively consistent across grades (9-16%) except at grade 1 (1%). Of consistently low frequency at all grades were: rating scale (0-2%), anecdotal notes (0-2%), journal (0-4%), and other (0-6%). The use of a few tools was related to grade: running records were found only at grades 1 and 2, journals were used only in grade 1, and checklists were most frequent in grade 1. In contrast, answer keys were infrequent at grades 1 (2%) and 2 (7%) and more frequent in grades 3-6 (16-23%).

Table 3 outlines the students' *tasks* corresponding to the assessment instances. Overall data show that almost a third (29%) involved a combination of tasks and that no single task is associated with more than 14% of assessments. Many of the low-frequency tasks were found also in combinations. For example, although listening was virtually never (0.4%) the main focus of an assessment, it was found more frequently in combinations. The relatively high frequency of tests could be related to the use of pre- and post-spelling tests.

The grade data on tasks showed few patterns, and because most overall frequencies were quite low, it is necessary to exercise caution interpreting any trends that appear. We grouped the data into three categories. The first includes those tasks that show no pattern by grade either because the range is narrow or the frequencies alternately dip and rise across grades: combination (range 25-32%), research report (3-12%), other (2-8%), oral work (1-3%), dramatic presentation (0-4%), discussion (0-1%), artwork (0-2%), listening (0-1%), and multimedia presentation (0-1%). The following showed patterns of increase by grade: test (0-23%), writing sample (6-16%), visual representation (1-10%), organizers (1-7%), and interview (0-3%). These increases are not necessarily regular, and again grade 3 appears to be an important place of change. For example, there are no assessments associated with tests at grade 1, only 8% at grade 2, and then a dramatic increase in grades 3-6. Decreasing trends were noted for the following: interactive reading/viewing (range 3-18%), reading/viewing (0-19%), science skill (0-5%), and worksheet (0-1%). Grade 3 again appeared as the period of shifting approaches. For example, assessment associated with interactive reading/viewing (e.g., guided reading) was much more frequent at grades 1 (16%) and 2 (18%) and then decreased dramatically (range 3-9%) for grades 3-6.

The *purposes* of the assessment instances were in large part dependent on teacher use. In 70% of cases we were unable to discern whether the assessment was summative, formative, diagnostic, or a combination of these. Thus in most instances, the programs do not give teachers directions on how to use the information gained from the assessment. The given purpose for an additional 22% of assessments was formative, and 8% were coded as having a combination of stated purposes. There were virtually no assessments for which the programs clearly identified a summative (.4%) or diagnostic (.1%) purpose. This overall pattern was replicated at each grade, although the actual percentages differed slightly.

Over half (55%) of the assessments we identified did not correspond to learning objectives either because no learning objective was given (30%) or

Table 3
Frequencies and Percentages of Assessment Tasks by Grade

| Grade 1 | Grade 2 | | Grade 3 | | Grade 4 | | Grade 5 | | Grade 6 | | Total | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Assessment Tasks* | *f* | *%* | *f* | *%* | *f* | *%* | *f* | *%* | *f* | *%* | *f* | *%* | *f* | *%* |
| Combination of Tasks | 35 | 32 | 28 | 25 | 35 | 28 | 36 | 30 | 35 | 30 | 34 | 29 | 203 | 29 |
| Test | 0 | 0 | 9 | 8 | 27 | 21 | 28 | 23 | 18 | 16 | 18 | 16 | 100 | 14 |
| Writing Sample | 7 | 6 | 10 | 9 | 16 | 13 | 14 | 11 | 19 | 16 | 13 | 11 | 79 | 11 |
| Interactive Reading/Viewing | 17 | 16 | 21 | 18 | 9 | 7 | 4 | 3 | 8 | 7 | 10 | 9 | 69 | 10 |
| Research Report | 7 | 6 | 14 | 12 | 5 | 4 | 7 | 6 | 3 | 3 | 8 | 7 | 44 | 6 |
| Visual Representation | 1 | 1 | 2 | 2 | 8 | 6 | 11 | 9 | 7 | 6 | 12 | 10 | 41 | 6 |
| Reading/Viewing | 21 | 19 | 13 | 11 | 0 | 0 | 5 | 4 | 1 | 1 | 0 | 0 | 40 | 6 |
| Other Task | 7 | 6 | 2 | 2 | 7 | 6 | 2 | 2 | 9 | 8 | 2 | 2 | 29 | 4 |
| Organizers | 1 | 1 | 3 | 3 | 5 | 4 | 4 | 3 | 6 | 5 | 8 | 7 | 27 | 4 |
| Science Skill | 4 | 4 | 5 | 4 | 6 | 5 | 0 | 0 | 2 | 2 | 2 | 2 | 19 | 3 |
| Oral Work | 3 | 3 | 3 | 3 | 3 | 2 | 1 | 1 | 4 | 3 | 4 | 3 | 18 | 3 |
| Dramatic Presentation | 4 | 4 | 2 | 2 | 2 | 2 | 3 | 2 | 0 | 0 | 3 | 3 | 14 | 2 |
| Student Planned Interview | 0 | 0 | 0 | 0 | 2 | 2 | 4 | 3 | 2 | 2 | 0 | 0 | 8 | 1 |
| Worksheet | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | .4 |
| Discussion | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 3 | .4 |
| Artwork | 1 | 1 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 3 | .4 |
| Listening | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 3 | .4 |
| Multimedia Presentation | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | .1 |
| Total per grade | 109 | 100 | 114 | 101 | 127 | 102 | 122 | 100 | 116 | 101 | 116 | 100 | 704 | 101 |

because coders determined that the assessment was not a good indicator of whether a student had achieved the given learning objective(s) (25%). This overall finding held for all grades with slight variation. At all grades the largest number of assessments (41-48%) was coded as corresponding to stated learning objectives. Except at grade 1, more assessments were unrelated to any learning objectives (26-33%) than assessments that were poor indicators of their corresponding learning objectives (20-33%).

The most frequently implicated *content area* for all assessments was language arts only, a finding that held both overall (41%) and at all grades (36-50%). Examples involving only language arts content would include instances of assessing students' sentence structure or the understanding and use of pronouns. Although all other assessments involved language arts content by virtue of students' involvement in reading, writing, listening, speaking, viewing, or representing, additional content areas were coded as detected in the instances. The next two most frequently occurring codes were for cross-curricular skills and content combinations that included science.

Assessments coded as containing cross-curricular content included directions that involve thinking processes and strategies that can apply to many subject areas such as using research skills, summarizing, reasoning, critical thinking, organizing, or participating and engaging in learning. For example, the following teacher direction shows assessment of participation in learning and on-task behaviour that is not tied to a particular subject area: "Consider: Does the child need to have instructions for the activity repeated? Does the child begin a task without unnecessary delay? Does the child stay on task until the work is completed?" (Ginn 2, *Round and Round*, p. 90). Also included were assessments that involve thinking processes and strategies that go beyond pure language arts content and involve higher-level thinking skills. For example, teachers are asked to consider the question, "Can the child identify details in this story that are similar to/different from details in others?" (Ginn 2, *Tales Near and Far*, p. 221). Because this assessment instance involved comparing and contrasting, skills that could be used in a variety of content areas, it was coded as containing cross-curricular content. Overall, 24% of assessments were given this code.

The third most frequent content code overall was for subject combinations that included science. Assessment instances given this code could involve any combination of subject areas (in addition to language arts) as long as science content was implicated in some way. For example, the following instance involves assessing children's understanding and recall of some life science content as well as some cross-curricular organization and presentation skills: "Listen as the children recall and record facts about the insects and birds from this unit. Are the children able to organize their thoughts and present them clearly? Are the facts relevant and accurate?" (Gage 1, p. 261). Overall, in the 233 instructional units covering the 238 selections that contain science content, only 14% of assessment instances were coded as content combinations including science. The patterns of occurrence for cross-curricular skills and combinations including science were almost identical across the grades. With the exception of grade 1, assessments involving cross-curricular skills were always second most frequent (range 21-29% in grades 2-6), and combinations including science were third most frequent (range 10-17% in grades 2-6). This order was reversed in grade 1 (13% and 17% respectively).

The results for the remaining content codes can be divided into two categories: those that appear with moderate frequency across the grades and those that are infrequent. The first category includes cases where content area cannot be discerned (4-10%), science (2-10%), interpersonal skills (2-6%), and other combinations not involving science (2-5%). Percentages for these four content codes were always lower than the first three outlined above, although not in the same order in each grade. Their order of overall frequency is as given above with percentages of 7%, 5%, 4%, and 3%. Of these, we were interested in assessments that implicate science content such as in the following example where students are asked to write explanations about the effectiveness of triangular shapes for making and building things: "Sample Criteria for Explanations: name what you are explaining; tell why it is used; tell how it is used; explain why it is important to us" (Nelson 3, *Hand in Hand*, p. 96). Because this task related to science, particularly technological problem-solving, which was one of the facets in our definition of science content, the content was coded as science (in addition to language arts). Thus across these selections containing science content, only 5% of assessments were coded as having science as their sole focus.

The remaining content areas occur infrequently across the grades (range 0-3%). Overall frequencies range from 0-1%, indicating that these content areas are rarely implicated in the main focus of an assessment although they may play a role in the combination options.

*Discussion*

Several aspects of our assessment findings can be viewed as supportive of scientific literacy goals. The findings on both *forms* and *tools* show that variety is present along these dimensions, which should be viewed positively given that position statements for both literacy (IRA & NCTE, 1994) and science education (NSTA, 2001) advocate multiple forms of assessment. There are some clear patterns of dominance, however, in that *observations* and *self/peer assessments* dominate forms and *guiding questions* dominate tools. This pattern may be viewed positively in that it suggests promotion of observation skills and a spirit of inquiry, essential elements in the scientific enterprise.

The dominant forms and tools often occur together: 76% of observations and 31% of self-assessments use guiding questions or prompts. For example, in a grade 2 instance of observation, the teacher was provided with the following questions: "Do children: record key observations accurately? recognize patterns and relationships in the charts and graphs? generate questions about their observations?" (Nelson 2, *Reach Out*, p. 137). Guiding questions or prompts were also provided for students as in the following self-evaluation on library research: "Do you: clearly understand what it is you want to research? make a list of questions you'd like to find answers to before you start? know how to locate resources on a topic? use a book's index and table of contents to see if it might have the information you're looking for?" (Gage 5, p. 49). Thus in many instances, teachers are given guidance on becoming good observers, and students are involved in the assessment process, a feature advocated by both literacy (Cohen & Wiener, 2003) and science educators (Coffey, 2003).

The findings on the purposes of the assessments also can be construed positively. Because the majority purpose was either formative or dependent on teacher use, there appears to be encouragement for teachers to use these assessments to gather information that would help students in their learning. The programs, therefore, give the appearance at least of emphasizing assessment for learning versus assessment of learning (Chappuis & Stiggins, 2002; Stiggins, 2002, 2005), and the value of this approach is well recognized in both science education and the literacy fields (Atkin, Black, & Coffey, 2001; Roscoe & Mrazek, 2005).

Other findings were less positive. Assessment was associated with a variety of student tasks, many in combination. Although this in itself may be a positive feature of the programs, the tasks most directly associated with science were quite infrequently associated with assessment. Our data show that only 3% of all assessments focused on tasks that involve students using skills or processes that are specific to the science domain (e.g., writing a scientific report, reporting observations, describing experiments). This finding may be due to a paucity of such tasks in the units. Although other tasks have aspects of relevance to the development of scientific literacy, their relationship to science is less direct. For example, a group task of researching a creature to write a story (Nelson 2, *Reach Out*, p. 161) may involve gathering information and learning some life science content that has scientific importance, but the task focus (writing a story) is of less relevance to the scientific enterprise. Assessment criteria for research presentations that include points like "follow the story model" suggest that teaching students to write scientific text is not a primary goal of the unit. The fact that the reading selection is a narrative affects both the subsequent task and assessment suggestions. Thus although there may be scientific value in this unit, the potential for teaching students to read and write scientific text, which tends to be expository and argumentative rather than narrative (Norris, Guilbert, Smith, Hakimelahi, & Phillips, 2005), is secondary and questionable.

The findings on the correspondence of the assessments to learning objectives and the content areas implicated in the instances are also not encouraging. Only about half (45%) of all assessments were clearly aligned with learning objectives, and these did not necessarily involve science. In addition, the content results show that most of the assessments were concerned with language arts content only. If we look at the percentage of assessments that have most relevance to science, we find the following: cross-curricular skills (24%), science (5%), and combinations including science (14%). Therefore, fewer than half (43%) of the assessments that corresponded with units containing scientific material had any clear connection to science content, and well over half of these were cross-curricular in nature. Thus whatever use these assessments may have for literacy learning, their relationship to the scientific realm is less than robust, and the relationships among objectives, instruction, and assessment are unlikely to be optimal for promoting scientific literacy in these programs.

Although the assessment findings noted here are suggestive, they do not lead directly to an evaluation of their potential usefulness. The value of the assessments is intimately related to the nature and quality of the student selections and the corresponding instructional suggestions. There is a complex relatedness among the given objectives, the student text, the instructional suggestions, and the assessments. Several examples may serve to illustrate. As a backdrop, consider two findings from our earlier study on the nature of the science texts and the corresponding instructional suggestions (Smith et al., 2006): (a) scientific material is found in a variety of text types and the quantity and quality of this material varies, and (b) the instructional guidance does not always capitalize on the fundamental scientific concepts that are in the text.

A first example involves the story *Ladybug Garden* by Godkin (1995). This selection, a narrative with a good deal of scientific content, is used by publishers in varied ways. An analysis of the assessments alone can be misleading as to the potential usefulness of the units. We found no assessments in one publisher's instructional unit (Gage, grade 3) although the other (Nelson, grade 4) contained four, two of which were labeled and involved science content or cross-curricular skills. An examination of the number and attributes of the assessments contained in the units would without consideration of other factors lead to a straightforward conclusion that the Nelson unit is superior to the Gage. However, this is not necessarily the case because the Gage unit provides instructional guidance that focuses more on the scientific concepts in the story and the Nelson unit focuses more or less on fact-finding. Although identifying facts in fiction is not an irrelevant exercise, the Gage approach is more conducive to teaching reading as inquiry, which we view as essential to fostering the fundamental sense of scientific literacy.

A second example concerns a Ginn grade 4 unit using an expository text *Dancing Bees* (modified from *The Big Bug Book*) by Facklam (1994), which has excellent potential for teaching reading in science. We identified five assessments in this unit, two of which were labeled and implicated science content in combination with cross-curricular skills. These assessments are aligned with a learning outcome stated "recount key ideas and information." The potential of the text or the assessments is, however, obscured with a plethora of instructional suggestions reflecting multiple purposes, many of which distract attention from the scientific concepts in the text. There are twice as many suggested curricular links to the arts (involving drawing a comic strip and performing a dance) than to science (gathering information about bees from a wide range of sources). The language arts activities include reader response suggestions such as discussion about how robots help people, imitation of the bees' waggle dance, looking at general features of nonfiction texts, spelling, and viewing/representing by sharing personal interpretations of dance directions. Three (unlabeled) assessments are associated with these activities. Although one suggested activity includes an organizer for noting important information, thus focusing some attention on fundamental scientific concepts in the text, opportunities are lost to teach students to read science text through interpreting scientific questions that seek explanations, interpreting descriptions of research methods and experimental tests, and distinguishing evidence from conclusions. Every one of these opportunities was present in the *Dancing Bees* selection. The selection includes a clear statement of the question that guided the scientific research described and of the motivation for that question. The selection also devotes considerable space to a detailed description of the observational methods used to gather data and of the experimental design employed. In addition, some of the key data are presented and conclusions based on those data advanced. In order to interpret this selection well, instructional techniques are required that draw attention to and explain to students these features of the text. As well, and relevant to the research described here, assessment techniques are required to learn how well students have learned to read text such as this. Unfortunately, neither the needed instructional nor assessment techniques were present. The assessments that were suggested may play a role in keeping some scientific information central to the reading experience, but even this may be lost if the teacher focuses elsewhere. Unless quality texts such as *Dancing Bees* are accompanied with aligned learning objectives, instructional guidance, and assessment suggestions that together optimize the scientific content, children's opportunities to learn to read and write scientific material will probably be compromised.

## *Conclusions and Implications*

This investigation examined the assessments connected with science content selections in widely used commercial reading programs. Interpreted in the context of a prior investigation into science content and corresponding reading instruction in these programs (Smith et al., 2006), the results suggest that there are serious limitations to what can be expected from use of these commercial reading programs as far as the fostering of scientific literacy is concerned. Although some aspects of the assessments included in science content units have relevance for the enterprise, their negative features outweigh the positive. To the extent that effective assessment guides instruction, the assessments in these units will not guide instruction that focuses on the fundamental scientific concepts that are in these texts. Thus although students may learn some generally useful literacy skills and strategies that may transfer to the scientific domain, their opportunities for learning to read and write scientific texts will be curtailed. These findings fortify the conclusion of our prior investigation that even the limited potential of these programs for fostering scientific literacy is unlikely to be realized.

The notion that science and literacy instruction should be integrated is increasingly popular. Although we emphatically support this idea, we emphasize the importance of giving careful consideration to the nature and purpose(s) of the integration. As Pratt and Pratt (2004) note, it is necessary to "be clear about the goals of the integration" rather than presenting the integration "as a goal in itself" (p. 395). It may be that some of the difficulties with using commercial reading programs for fostering scientific literacy hinge on lack of clarity in integrating goals.

Although the inclusion of scientific content in commercial reading programs may be generally regarded as a positive feature of the programs and convey an impression of usefulness for teaching students to read scientific texts, the goals of the integration may be inconsistent with the perceptions of users of the programs. The inclusion of informational texts may be intended primarily to provide students with opportunities to encounter a variety of text types although the teaching of science content from other science texts is assumed on the part of publishers. Because of limitations inherent in programs that focus on teaching literacy skills and strategies apart from a sustained and in-depth exploration of content (Smith et al., 2006; Walsh, 2003), the well-intentioned integration of scientific material into these programs may have limited usefulness for developing scientific literacy.

*References*

Alvermann, D.E., & Boothby, P.R. (1982). Text differences: Children's perceptions at the transition stage in reading. *Reading Teacher, 36*, 298-302.

Atkin, J.M., Black, P., & Coffey, J. (Eds.). (2001). *Classroom assessment and the National Science Education Standards*. Washington, DC: National Academy Press.

Brickhouse, N. (1990). Teachers' beliefs about the nature of science and their relationship to classroom practice. *Journal of Research and Development in Education, 15*(4), 13-18.

Chappuis, S., & Stiggins, R. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40-43.

Coffey, J.E. (2003). Involving students in assessment. In J.M. Atkin & J. Coffey (Eds.), *Everyday assessment in the science classroom* (pp. 75-87). Arlington, VA: National Science Teachers Association.

Cohen, J.H., & Wiener, R.B. (2003). *Literacy portfolios: Improving assessment, teaching, and learning* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

*Collections.* (1996, 1997, 1998, 1999, 2000). Scarborough, ON: Prentice Hall Ginn.

*Cornerstones Canadian Language Arts.* (1998, 1999, 2000, 2001). Toronto, ON: Gage Educational.

Craig, M.T., & Yore, L.D. (1996). Middle school students' awareness of strategies for resolving comprehension difficulties in science reading. *Journal of Research and Development in Education, 29*, 226-238.

Facklam, M. (1994). *Big bug book*. Boston, MA: Little Brown.

Flood, J., & Lapp, D. (1987). Forms of discourse in basal readers. *Elementary School Journal, 87*, 299-306.

Godkin, C. (1995). *Ladybug garden*. Markham, ON: Fitzhenry & Whiteside.

Graesser, A.C., McNamara, D.S., & Louwerse, M.M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A.P.Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82-98). New York: Guilford Press.

Harlen, W. (1997). Primary teachers' understanding in science and its impact in the classroom. *Research in Science Education, 27*, 323-337.

Heselden, R., & Staples, R. (2002). Science teaching and literacy, Part 2: Reading. *School Science Review, 83*, 51-62.

IRA & NCTE. (1994). *Standards for the assessment of reading and writing*. Newark, DE: International Reading Association and National Council of Teachers of English Joint Task Force on Assessment.

Johnston, P., & Costello, P. (2005). Principles for literacy assessment. *Reading Research Quarterly, 40*, 256-267.

Levitt, K.E. (2001). An analysis of elementary teachers' beliefs regarding the teaching and learning of science. *Science Education, 86*, 1-22.

Morrow, L.M., & Gambrell, L.B. (2000). Literature-based reading instruction. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (vol. 3, pp. 563-586). Mahwah, NJ: Erlbaum.

Moss, B., & Newton, E. (2002). An examination of the informational text genre in basal readers. *Reading Psychology, 23*, 1-13.

Murphy, S. (1991). Authorship and discourse types in Canadian basal reading programs. *Reflections on Canadian Literacy, 9*(3, 4), 133-138.

National Science Teachers Association. (2001). *Position statement: Assessment*. Retrieved February 21, 2006, from: http://www.nsta.org/positionstatement&psid=40

*Nelson Language Arts*. (1998, 1999, 2000, 2001). Scarborough, ON: Nelson Thomson Learning.

Neuendorf, K.A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage.

Newton, D.P., & Newton, L.D. (2000). Do teachers support causal understanding through their discourse when teaching primary science? *British Educational Research Journal, 26*, 601-613.

Norris, S.P, Guilbert, S.M., Smith, M.L., Hakimelahi, S., & Phillips, L.M. (2005). A theoretical framework for narrative explanation in science. *Science Education, 89*, 535-563.

Norris, S.P., & Phillips, L.M. (1994). Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching, 31*, 947-967.

Norris, S.P., Phillips, L.M., Smith, M.L., Guilbert, S.M., Stange, D.M., Baker, J.J., & Weber, A.C. (2008). Learning to read scientific text: Do elementary school commercial reading programs help? *Science Education, 9*, 765-798.

Otero, J.C., & Companario, J.M. (1990). Comprehension evaluation and regulation in learning from science tests. *Journal of Research in Science Teaching, 27*, 447-460.

Pajares, M.F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307-322.

Palinscar, A.S., & Magnusson, S.J. (2001). The interplay of first-hand and second-hand investigations to model and support the development of scientific knowledge and reasoning. In S.M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 151-187). Mahwah, NJ: Erlbaum.

Pappas, C.C. (2006). The informational book genre: Its role in integrated science literacy research and practice. *Reading Research Quarterly, 41*, 226-250.

Phillips, L.M., & Norris, S.P. (1999). Interpreting popular reports of science: What happens when the reader's world meets the world on paper? *International Journal of Science Education, 21*, 317-327.

Phillips, L.M., & Norris, S.P. (2003). Middle grade readers' ability to interpret pragmatic meaning. In G. Shiel & U. Ní Dhálaigh (Eds.), *Other ways of seeing: Diversity in language and literacy* (vol. 1, pp. 223-230). Dublin: Reading Association of Ireland.

Phillips, L.M., Smith, M.L., & Norris, S.P. (2005). Commercial reading programs: What's replacing narrative? In B. Maloch, J.V. Hoffman, D.L. Schallert, C.M. Fairbanks, J. Worthy, (Eds.). *54th yearbook of the National Reading Conference* (pp. 286-300). Oak Creek, WI: National Reading Conference.

Pomeroy, D. (1993). Implications of teachers' beliefs about the nature of science: Comparison of the beliefs of scientists, secondary science teachers, and elementary teachers. *Science Education, 77*, 261-278.

Pratt, H., & Pratt, N. (2004). Integrating science and literacy instruction with a common goal of learning science content. In E.W. Saul (Ed.), *Crossing borders in literacy and science instruction* (pp. 395-405). Arlington, VA: International Reading Association and National Science Teachers Association.

Roscoe, K., & Mrazek, R. (2005). *Scientific literacy for Canadian students: Curriculum, instruction and assessment*. Calgary, AB: Detselig.

Rowell, P.M., & Ebbers, M. (2004). School science constrained: Print experiences in two elementary classrooms. *Teaching and Teacher Education, 20*, 217-230.

Saenz, L.M., & Fuchs, L.S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education, 23*(1), 31-41.

Saul, E.W. (Ed.). (2004). *Crossing borders in literacy and science instruction: Perspectives on theory and practice*. Arlington, VA: International Reading Association and National Science Teachers Association.

Shanahan, C. (2004). Better textbooks, better readers and writers. In E.W. Saul (Ed.), *Crossing borders in literacy and science instruction: Perspectives on theory and practice* (pp. 370-382). Arlington, VA: International Reading Association and National Science Teachers Association.

Shymansky, J.A., Yore, L.D., & Good, R. (1991). Elementary school teachers' beliefs about and perceptions of elementary school science, science reading, science textbooks, and supportive instructional factors. *Journal of Research in Science Teaching, 28*, 437-454.

Smith, M.L., Phillips, L.M., Leithead, M.R., & Norris, S.P. (2004). Story and illustration reconstituted: Children's literature in Canadian reading programs. *Alberta Journal of Educational Research, 50*, 391-410.

Smith, M.L., Phillips, L.M., Norris, S.P., Guilbert, S.L., & Stange, D.L. (2006). Scientific literacy and commercial reading programs: An analysis of text and instructional guidance. In J.V. Hoffman, D.L. Schallert, C.M. Fairbanks, J. Worthy, & B. Maloch (Eds.), *55th yearbook of the National Reading Conference* (pp. 293-308). Oak Creek, WI: National Reading Conference.

Stiggins, R.J. (2002). Assessment crisis: The absence of assessment FOR learning. *Phi Delta Kappan, 83*, 758-765.

Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*, 324-328.

Trumper, R. (2003). The need for change in elementary school teacher training: A cross-college age study of future teachers' conceptions of basic astronomy concepts. *Teaching and Teacher Education, 19*, 309-323.

Vanderwolf, C.H., Cook, M., Coutts, R.T., & Cropp, D. (2005). *Teaching science in the 21st century: An examination of Canadian science curricula from kindergarten to grade 12*. Waterloo, ON: Society for Quality Education

Walsh, K. (2003, Spring). Basal readers: The lost opportunity to build the knowledge that propels comprehension. *American Educator*, 24-27.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education.* Philadelphia, PA: Open University Press.