From **Gal, I. & Garfield, J. B. (editors).** **The Assessment Challenge in Statistics Education**. IOS Press, 1997 (on behalf of the ISI). Pages 107-121. ISBN 90 5199 333 1. Copyright holder: International Statistical Institute. Permission granted by ISI to the IASE to make the chapters from this book freely available on the Internet. The References section is a separate chapter and is available on the Internet at http://www.stat.auckland.ac.nz/~iase/publications/assessbkref.

9. Assessing Statistical Thinking Using the Media

Jane M. Watson, University of Tasmania

Purpose

The goals of this chapter are (a) to address the need to assess statistical thinking as it occurs in social settings outside the classroom, (b) to suggest a hierarchy for judging outcomes, (c) to provide examples of viable assessment based on items from the media, and (d) to discuss the implications for classroom practice.

INTRODUCTION

At the end of the nineteenth century H. G. Wells claimed that “statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write” (quoted in Castles, 1992, p. v). The growing emphasis, in recent years, on probability and statistics in the mathematics curriculum might be seen as acknowledging Wells’s prophecy. This increased statistical content is evident in the curricular documents of many western countries: for example, in Australia the mathematics curriculum includes *chance and data*, in England and Wales, *data handling*, in the United States, *statistics and probability*, and in New Zealand, *statistics*.

Beyond merely placing technical statistical topics in the curriculum, Wells’s emphasis on the needs of society in relation to statistical thinking is also reflected in curricular statements. In Australia it is found in *A National Statement on Mathematics for Australian Schools* (Australian Education Council [AEC], 1991) under headings which require students to “understand and explain social uses of chance” (p.175) and “understand the impact of statistics on daily life” (p.178). In the United States, the National Council of Teachers of Mathematics’ (NCTM) *Curriculum and Evaluation Standards* (1989) exhibits the same sentiments. There it is claimed that the curriculum should provide situations so students can develop an appreciation for “statistical methods as powerful means for decision-making” (p.105) and “the pervasive use of probability in the real world” (p.109). Hence probability and statistics are suggested for inclusion in the mathematics curriculum not only because of their innate worth as intellectual topics but also because of their application in dealing with issues in wider society.

The need for statistical thinking in social decision-making is exemplified every day in the news media, where reports appear on topics as wide-ranging as politics, health, town planning, environmental control, unemployment, sport, science, and attendance at cultural events. If evidence of the need for statistical literacy is found in the media, then the media is also an ideal vehicle to provide initial motivation for the study of statistics, applications of specific topics in the curriculum during instruction, and items for assessment in the final stages of learning. It is the contention of this chapter that the validity of context-based instruction, in this case using the media, is only ensured if the context is employed at all stages of the learning cycle.

The issues related to the assessment of statistical thinking are not only the province of mathematics teachers. As other curriculum areas acknowledge the importance of the analysis of data and risk in relation to their subject matter, the need for statistical literacy is seen in many differing contexts. Recent curricular reform in Australia illustrates this trend. *A Statement on Studies of Society and Environment for Australian Schools* (AEC, 1994a) includes a process strand, "Investigation, Communication and Participation," which states the following expectations for students:

> Students gradually build up the skills involved in research, processing data and in interpreting or applying findings.... This is a foundation for predicting possible solutions to the problem, constructing hypotheses, considering other approaches to inquiry, and designing suitable methods for gathering and organising information. Sources of information are assessed for their authenticity and credibility (p.11).

These expectations rely very heavily on skills of statistical thinking, and associated specific outcomes are listed in the accompanying *Profile* document (AEC, 1994b). Similar statements are found in curriculum documents in science, health and physical education, and technology. Teachers of many subjects in the school curriculum are facing the task of teaching and assessing statistical thinking in some context.

TARGET SKILLS AND LEVELS OF ACHIEVEMENT

Once we accept that statistical thinking in social contexts is an important part of statistical education, it is necessary to describe the associated skills and their levels of complexity. This will assist teachers in structuring learning experiences and planning the related assessment. The skills required to interpret stochastic information presented in society, often in the form of media reports, can be represented in a three-tiered hierarchy: (a) a basic understanding of probabilistic and statistical terminology, (b) an understanding of probabilistic and statistical language and concepts when they are embedded in the context of wider social discussion, and (c) a questioning attitude which can apply more sophisticated concepts to contradict claims made without proper statistical foundation. These skills represent increasingly sophisticated thinking and are consistent with models of learning from developmental psychology (see for example, Biggs & Collis, 1982, 1991; Case, 1985; Watson, Collis, Callingham & Moritz, in press). Each will be considered in turn.

Tier 1: Basic understanding of terminology

At the first stage of the hierarchy there are the skills related directly to specific topics in the curriculum; these are generally taught in a conventional fashion with students creating and analyzing their own data sets. At various levels of the curriculum the topics include percentage, median, mean, specific probabilities, odds, graphing, measures of spread, and exploratory data analysis. All can be taught without reference to social issues, and detailed discussions of their teaching and assessment are found elsewhere in this volume. If, however, students ever express the query, “Why do we have to learn this?”, the evidence from the media is very easy to find. In fact, asking students to find it for themselves may be the most motivating avenue to pursue. If used at this level media extracts should be very straightforward. Searches are likely to produce examples such as the following short story (“Weight guidelines,” 1995), which mentions percentage and average.

> United States federal guidelines on healthy weights for men and women are too lenient and may be encouraging Americans to weigh too much according to a new study. The study found that women of average weight in the US had a 50 per cent higher chance of heart attack than did women weighing 15 per cent below average.

Although used initially to motivate basic definitions, discussion can lead to the next level of the skills hierarchy.

Tier 2: Embedding of language and concepts in a wider context

Once students with some rudimentary statistical concepts in hand are exposed to the media, a second need--to read and interpret written reports, rather than just perform computations—becomes important. Some students who have excelled in the traditional symbolic aspects of the mathematics curriculum resist the requirement for reading, interpreting, and writing when mathematics is presented in non-symbolic contexts. In all aspects of mathematical thinking, however, the need for application, interpretation, and communication skills is being recognized (e.g., Schoenfeld, 1992). The specific necessity to tie statistical and literacy skills together is acknowledged in curriculum documents around the world; for example, in Australia the language skills for young children are prescribed in two outcomes for the early and late elementary years (AEC, 1991): “use with clarity, everyday language associated with chance events” (p. 166), and “make statements about how likely are everyday experiences which involve some elements of chance and understand the terms ‘chance’ and ‘probability’ in common usage” (p. 170). These objectives demonstrate the need to begin early in tying probability and statistics to everyday experience.

As students mature, the need for language interpretation skills is no less important and the media readily supply examples. Newspaper headlines could provide a motivation for the need to relate likelihood to everyday events, a basis for class discussion of relative likelihood, or an assessment task. Students could be asked to rank the following five headlines from least likely to most likely, with justification: (a) “Willis looks sure bet for Treasurer,” (b) “Freak proves it’s no fluke,” (c) “Blacks’ shocking jail odds,” (d) “Stingrays grab second chance,” and (e) “Ghost of a chance something funny will happen.”

The understanding of risk is important in many decision-making situations and other curriculum areas which demand an appreciation of risk analysis. This involves making assumptions of student understanding based on work in probability in the mathematics class. Consider, for example, the following extract (Ewing, 1994) that might be used in a mathematics class or a science class.

> There is a one-in-10,000 chance that an asteroid or comet, more than two kilometres in diameter, will collide with Earth in the next century, killing a large proportion of the population, according to space scientists.... An article in the British journal *Nature* says the risk is great enough to justify a space surveillance system that would warn scientists of approaching objects and allow them to deflect them with nuclear explosions.... A person living in the United States has a greater chance of dying from a comet impact (one in 20,000) than being killed in a flood (one in 30,000) or a one-in-100,000 risk of death from a venomous bite or sting.

Again the need is seen for an understanding of probabilistic language if sense it to be made of the article.

Statistical language offers the same challenge when placed in context. The extract on weight noted earlier also can be used at this level to reinforce understanding. At this second level of sophistication, more than just basic definitions are needed to be successful. It is necessary to recognize these in other contexts and be able to make sense of claims which are made. Since on most occasions statistics are presented correctly, the requirement at this level is to understand and interpret statistics in order to draw conclusions and make decisions.

Tier 3: Questioning of claims

At the highest level of the statistical thinking hierarchy, students possess the confidence to challenge what they read in the media. It sometimes happens that claims are made without proper statistical foundation, either inadvertently or purposefully. Whether there is an intention to mislead or just insufficient information, students need to be made aware of the expectation that they must constantly question conclusions. The specific questioning skills required at this level are exemplified by Gal (1994) and relate to sampling, the distribution of raw data, appropriate use of statistics, graphs, causal claims made, and probabilistic statements.

The astronomical example given earlier provides more than just an opportunity to interpret risk in terms of probability. It gives the opportunity to develop (or assess) student skills in interpreting stochastic information and questioning the motives of people who use it. The purpose of the article was to convince politicians or finance-granting authorities to fund a $67 million-plus project to list all potentially threatening asteroids large enough to precipitate a global catastrophe. Questions of how the claimed probabilities were obtained are not answered in the article and the statistically literate reader would be highly suspicious of the figures given the desire of the scientists to fund their pet project. Thus, to move students from a situation where they automatically believe everything they read in the media to one where they intelligently question data and claims is an important aspect of statistical literacy.

Similar skills are important in relation to an understanding of samples and populations and their use in media reports. Small samples, samples without a mention of their size, and non-representative samples should all be treated with suspicion and results based on them considered with scepticism. An extract like the following ("Decriminalise," 1992) can be used to elicit higher order analysis skills from students in this area.

> Some 96 percent of callers to youth radio station Triple J have said marijuana use should be decriminalised in Australia. The phone-in listener poll, which closed yesterday, showed 9,924—out of the 10,000-plus callers—favoured decriminalisation, the station said. Only 389 believed possession of

> the drug should remain a criminal offence. Many callers stressed they did not smoke marijuana but still believed in decriminalising its use, a Triple J statement said.

Although the sample size is very large, the question of the population represented by the listening audience and the voluntary nature of the phone-in procedure lead to scepticism about the claims that could be made on the basis of the report.

For classroom purposes it is important to appreciate the increasingly complex nature of the thinking involved as students move from developing a basic understanding of statistical terminology and concepts in a mathematical context, to understanding and applying them in a wider social context, to questioning their use by those who may wish to mislead members of society. Assessment of these skills goes hand-in-hand with their teaching and learning. There is no reason why the media cannot be used as a basis for assessment as it may have been earlier for initial motivation or classroom discussion in conjunction with learning concepts.

BASIC ISSUES IN ASSESSMENT OF STATISTICAL THINKING

Although there can be many purposes for assessing statistical thinking, including research or the development of state- or nation-wide norms, the emphasis here is on assessment to inform teachers for instructional purposes and students for progress reports. It is hence the formative or summative nature of the assessment which is of interest. Even with these narrower aims in mind, however, the scope for using innovative methods of assessment is wide (see e.g., Webb, 1993).

For classroom teachers the major result of any type of assessment should be the evaluation of whether unit objectives have been met and the informing of future instruction, be it for remedial work following a test or for planning next year's teaching sequence. In this context the types of media-based items introduced in the next section can provide formative evaluation of the teaching program. Such items may be administered in testing situations or used for group discussion. In either situation it is the level of response achieved which should be the concern of the teacher. Can students interpret statistical information in context and can they question suspicious claims? If not, then it will be necessary to adapt teaching practice to assist students in achieving higher levels.

A secondary necessity for teachers may be to provide a summative measure of students' performances. For this purpose schemes such as those suggested for problem solving in the NCTM *Standards* (1989) or by Charles, Lester and O'Daffer (1987) may be modified for use. This involves assigning increasing integer values as students achieve higher levels of sophistication in their responses. This approach will be outlined briefly in relation to the items presented in the next section.

It will be seen that the items introduced here are of the *open-ended* type that require students to provide written (or oral) responses rather than to choose from a selection of teacher-composed alternatives. This method is used with the intent of allowing students to demonstrate statistical understanding and questioning ability which would not be possible in a multiple-choice format. This is consistent with Schoenfeld's (1992) comments in terms of the assessment of mathematical thinking. In supporting the use of items of the type suggested here he says,

> To state the case bluntly, current assessment measures (especially the standardized multiple-choice tests favored by many administrators for "accountability") deal with only a minuscule portion of the skills and perspectives encompassed by the phrase *mathematical power*... (p. 365).

At the third tier of statistical thinking assessed here, indeed a high degree of mathematical power is being demonstrated.

It will be noticed also that computation is not required as part of the assessment used here; this is generally the case when media reports are used for assessment. There are two reasons for this. First, the media seldom provide raw data upon which computations can be performed. Second, the interpretative skills being assessed generally involve the understanding of concepts rather than computation. The purpose of assessing interpretive skills is to discover if students can move to higher levels of cognitive functioning than are generally required to perform computations. The skills here encompass the communication of ideas as well as the recognition of meaning in context (NCTM, 1991).

There is little doubt that it is easier to make decisions about calculations than about many of the responses shown in this chapter. The requirement to judge levels of sophistication in statistical thinking means that teachers must understand the concepts well enough not to be fooled by students who may be grasping at straws to fill in a blank space on a piece of paper. There has been some concern by mathematics educators (e.g., McGregor, 1993) that once writing about mathematics becomes part of assessment in the subject, teachers are likely to reward quantity rather than quality. Students who write many lines, very neatly, are apt to be given more marks than those who are succinct and perhaps not as tidy. Therefore, a big responsibility falls on the teacher to develop interpretive skills to a high level as well.

FORMATIVE ASSESSMENT WITH MEDIA ITEMS

In this section two examples will be presented to indicate how articles from newspapers can be used to assess students' interpretive statistical thinking. The items were used initially as part of a large survey of student understanding of statistics presented in social contexts which was carried out with 670 students in Grades 6 and 9 in Tasmania. The responses of some of these students will be used as a basis for discussion of the items presented here. The responses will illustrate the variety and level of understanding shown by students in these grades. The students who were administered the Media Survey had had mixed classroom experiences with probability and statistics. Some of the Grade 6 students had been exposed to lessons on experimental probability, and some Grade 9 students had had a unit on statistics. The variety of responses provide a range from what could be described as naïve common sense to more sophisticated interpretive skills. In some cases intuition is used; it may or may not lead to acceptable outcomes.

The examples represent two content areas from the statistics curriculum, each presented in a different context, and offer the opportunity for students to question a contention made in each report and reach the third tier of the statistical thinking hierarchy. The content areas represented are graphical representation, using a pie chart in an economic context, and sampling, in a social science context. The use of media extracts related to probability is considered elsewhere (Watson, 1993).

There is no reason why these items cannot be used in teaching as well as assessment. The emphasis here, however, is on the type of achievement which can be expected when the media extracts are used for assessment. Webb (1992) suggests four components of the assessment process which are helpful in following through with the procedure: situation, response, analysis and interpretation. These will be used in the following examples to consider the formative assessment possible in relation to statistical thinking. *Situation* will be interpreted here to mean the context set by the question asked. The physical situation associated with the administration of the item may vary from classroom discussion, to group work, to individual written responses and will not be discussed for each example. The *response* phase will be illustrated with responses from the Tasmanian survey. The *analysis* will be based on levels associated with the latter two tiers of the statistical thinking hierarchy introduced earlier. Finally, *interpretation* will connect the analysis back to classroom practice.

The framing of questions seeking to elicit understanding of media extracts is not an easy task. The objective is to allow for various interpretations but perhaps to guide at some stage of the questioning protocol. Various forms of questions were trialed in the pilot work leading to the items used in this chapter. Those presented here may not be considered perfect, but they did produce a wide range of responses. Teachers should be aware of the necessity to think carefully about the questions they ask in order not to bias the responses.

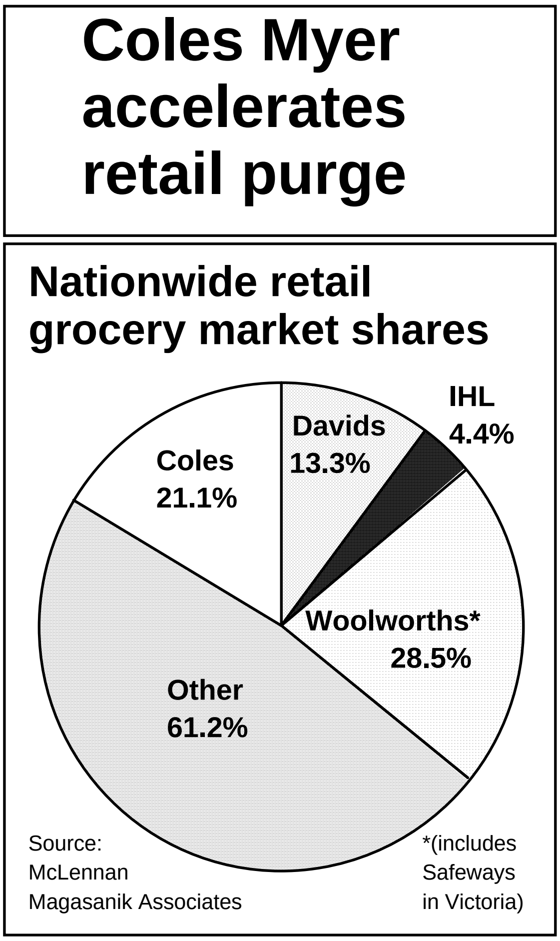
Graphical representation

Representation of data graphically is one aspect of statistics which has been in the mathematics curriculum for a long time and it might be expected that students are familiar with various types of representation. The work of Pereira-Mendoza and Mellor (1991) and Curcio (1987) reveals, however, that students are likely to have beliefs about the features of graphs that are different from what is expected. It is important to reinforce the understanding of graphing in the high school years; and as the following item shows, there is scope for testing the highest level of statistical thinking: the questioning of information presented.

Situation

The headline and pie chart shown in Figure 1 appeared, with an accompanying article, on page 1 of the *Australian Financial Review* (Webb, O'Meara & Brown, 1993). The questions in Figure 1 ask the meaning of the pie chart and whether there is anything unusual about it. It is assumed that the Tasmanian responses are similar to those that would be obtained in most developed countries.

It is obvious to those who understand the principle that a pie chart represents 100% of the quantity under consideration that there is something wrong with the pie chart in the figure. It is of great interest whether students will notice this mistake.



Explain the meaning of this pie chart.

Is there anything unusual about it?

Figure 1: Pie chart question

Response

Of Grade 6 students, there are quite a few who say they do not understand the pie chart and leave the question blank. This type of response diminishes greatly by Grade 9. The following responses would likely be acceptable to most teachers for explaining the meaning of the pie chart: (a) “It tells us what % they have sold,” (b) “That Coles got 21.1%, Davids 13.3%, Woolworths 28.5%, other 61.2%,” and (c) “It shows who out of the 5 markets who [sic] has the most share of the grocery market shares.” These responses have achieved the second tier of the statistical thinking hierarchy.

When open-ended responses are allowed, however, many are found to be lacking in some respect. The response, “It says how well those shops are selling,” has interpreted the situation correctly but has it contained enough information about the percentage shares of the market? The response, “It shows Nation wide retail grocery market shares,” represents a tautology, merely repeating the information in the title of the pie chart. The response, “It means how many people go there,” equates market share with number of shoppers. Is this a reasonable assumption? It also alludes to quantity (“how many”) rather than percentage. “I think it is about all the people who have taken out shares” misunderstands the term *share* in the context and illustrates the need to appreciate the language of the application before being able to make adequate statistical interpretations . Whereas the response, “How much money is being made,” has interpreted the context correctly, it has not mentioned the relative information for the different groups in the pie chart and again emphasizes quantity rather than percentage. Is this student just being careless or is the understanding not present?

In a classroom it may be possible to answer these questions directly—a big advantage over a large-scale impersonal survey. If the answers were part of a whole-class discussion of the pie chart’s meaning, they would probably be greeted with an encouraging, “Yes, that’s right. Now what else can we tell from the chart?” How the answers should be handled in a written format is more difficult.

Moving to the second part of the question in Figure 1, the recognition of the error in the pie chart represents a higher level of functioning: the third tier of the statistical thinking hierarchy. Those who can answer the second part correctly can give concise functional responses to the first part. Not all students who can answer the first part adequately, however, can produce the higher level response needed for the second part.

When they are asked if there is anything unusual in the pie chart many students give “No” responses, and many others fail to answer the question at all. Instead of recognizing the percentage error in the pie chart, students’ lower-level responses suggest instead that there is something “unusual” within the chart itself. The first five responses represent aspects not related to the statistical task in hand: (a) “It’s cut into all different shapes,” (b) “It doesn’t say what ‘Other’ shops are,” (c) “They are all decimal like 21.2, 13.3, 28.5, 61.2,” (d) “The black part,” and (e) “I can’t figure out why Woolworths have a star and the rest don’t.” The next three responses mention aspects which are related to the statistical representation but these are not the significant unusual feature to one who understands fully the creation of a pie chart: (a) “Other is bigger than the rest,” (b) “Coles is one of the smallest market shares,” and (c) “The heading doesn’t fit in.” The “other is bigger” response is the most common answer of this type. At the level at which this part of the question is directed, these responses would be considered not to have engaged the question; that is, they do not achieve a level of functioning higher than that required for the first part of the question.

The recognition of the percentage error in the pie chart can be noted in two ways: (a) “The percentages add up to 128.5. They should equal 100!!” and (b) “Where it has Other, it says 61.2% and the percentage of that section on the pie is less than 50%.” It could be argued that response (a) is more sophisticated in its realization of the incorrect total. Response (b) may not be accompanied by this realization but be the result of perceived inaccuracy in drawing the sections of the pie.

Analysis

Whereas on first glance, the question about the meaning of the pie chart may appear to be a relatively straight-forward application of this form of graphical representation in a particular context, the responses above show that partial understanding is possible from several perspectives. Hence to achieve a response which fully satisfies the first two tiers of the statistical thinking hierarchy—to show a basic understanding of the statistical representation and appreciate how that representation is embedded in a wider context—requires a fairly high level ability to relate ideas together. Intermediate level responses are exhibited above and it is possible to suggest an ordering as follows.

- No engagement with the item:

"Don't know."
"I was absent when we learned pie graphs."

- Single facet of the item accessed in response:

"That others is all the other grocery market."
"I think it is about all the people who have taken out shares."

- Appreciation of some more complex aspects of the pie chart representation:

"It says how well those shops are selling."
"It shows the percentage of people in the population who have taken out shares..."

- A full understanding which relates pie chart information in context:

"It tells us what % they have sold."
"That Coles got 21.1%, Davids 13.3%, ..."
"It shows who out of the 5 markets has the most share of the grocery market shares."

Students responding at this last level could be claimed to have satisfied the second tier of the hierarchy of applying statistical understanding in a social context. Reaching the third tier is reserved for those who could go further and question the percentage figures given in the pie chart. There is therefore a final level of achievement for this item.

- An ability to correctly question claims made:

"The percentages add up to 128.5. They should equal 100!!"
"Where it has Other, it says 61.2% and the percentage of that section on the pie is less than 50%."

The importance of ranking intermediate responses is to recognize the need to raise levels of response if given classroom opportunities and to provide a basis for giving summative feedback if required. To merely classify responses as *wrong* or *right* may mean that chances are missed to encourage progression to higher levels.

Interpretation

It is important for teachers to appreciate the degree of complexity involved in the intermediate responses which lead up to what would be considered a fully adequate reply. It is the acknowledgment of this complexity and the rewarding of movement from one level to the next that is part of the contribution a teacher can make to the improvement of performance by students. A class discussion which leads students from single facet replies to more complex responses representative of the second and third tier will assist students to form the connections to be able to produce higher level responses.

There are teachers who would claim that the graphical representation created by a pie chart is not fully understood until the second part of the item in Figure 1 is answered correctly. In that case the importance of including in the assessment an error which embodies the essence of the pie chart concept is immediately seen. Those who answer the first part adequately have shown an appreciation of some of the features of a pie chart. The necessity to represent a whole with 100% was not a feature mentioned by many in answer to the first question, but it is encouraging that many did recognize the lack of the property when asked to explore further. The second part of the pie chart item illustrates the importance of the third skill in the hierarchy: a questioning attitude which can apply the more sophisticated concepts to contradict claims made without proper statistical foundation.

Sampling

The importance of sampling as part of the statistics curriculum has been recognized more recently than graphing; it is now specifically mentioned, however, in current curriculum documents (e.g., AEC, 1991; NCTM, 1989). Sampling activities (e.g., Friel & Corwin, 1990; Landwehr, Swift & Watkins, 1987) are now suggested for both primary and secondary students, and the importance of using an appropriate sampling procedure is stressed in relation to instances of misleading claims being made (e.g., Watson, 1992). Deficiencies in students' basic understanding of sampling are also being documented (e.g., Rubin, Bruce & Tenney, 1991), which further supports the need for challenging assessment items in the area.

Situation

The extract shown in Figure 2 appeared in the "That's Life" (1993) column of the *Hobart Mercury*. The first part of the question associated with the report (see Figure 3) asks students for any criticisms they may have of the claims in the article. It is meant to be open-ended to allow for other than statistical criticisms if students focus on them. The second part provides a more specific opportunity to consider the handgun claim for the whole of the United States based on a sample from Chicago.

The concept to be assessed is the relationship between a sample and a population. As neither of these terms is mentioned in the article, students must have both the basic understanding of the terms and the ability to recognize them in a social context when other words, such as *poll,* are used. The essence of the questions, however, is to gauge if students have reached the third tier of statistical thinking where they can question the claim about an inappropriate population based on the sample. Because of this desire to assess the highest level of statistical thinking, it was necessary to word the questions in a way to avoid the words *sample* and *population*.

ABOUT six in ten United States high school students say they could get a handgun in they wanted one, a third of them within an hour, a survey shows. The poll of 2508 junior and senior high school students in Chicago also found 15 per cent had acutally carried a handgun within the past 30 days, with 4 per cent taking one to school.

Figure 2. Newspaper article about sampling.

Would you make any criticisms of the claims in this article?

If you were a high school teacher, would this report make you refuse a job offer somewhere else in the United States, say Colorado or Arizona?
Why or why not?

Figure 3: Questions for students

Response

Many students in both Grades 6 and 9 say “No” to both questions. Of the Grade 6 students in the Media Survey who express criticisms in answer to the first part of the question, none focus on the sample-population question. This may not be surprising since most Grade 6 students would not have been exposed to very sophisticated work on sampling. It is interesting, however, to note that the criticisms are mainly aimed at the implications of the claims, and not the content of the claims: (a) “Rules in USA schools are terrible!” (b) “They shouldn’t have hand guns at all,” and (c) “I think that the owners of the gun shops should be fined for letting the kids buy them.” Few Grade 6 students in any way question the report. The criticism, “I doubt that they could get one within an hour...”, is probably related to the student’s social experience, which would be very different from Chicago. The student who responds, “Because they could be lieing [lying],” is questioning the veracity of the evidence given by the Chicago students, which is indeed an issue when conducting surveys.

Many Grade 9 students make comments similar to Grade 6, but there are some who do appreciate the sample-population problem: (a) “Yes that with only 2508 students interviewed they could have only found [them] in [the] rough part of town,” (b) “They only interviewed people in Chicago,” (c) “The articles poll is done in a city with crime not in a country where the % could be 1,” and (d) “They are saying 6 in 10 United States high school students. But actually it is 6 in 10 Chicago high school students.” The first two responses focus on the sample without reference to the population but in each case it appears that the relationship is implicitly understood. The second two responses explicitly state the difficulty in the article.

The second question in Figure 3, related to teaching in another place than Chicago, offers a further chance for students to pick up the error in the article by suggesting a scenario which explicitly states a different geographical location for comparison. This added probe results in some Grade 6 students noticing that the sample from Chicago may not be representative of Colorado or Arizona, even though they had originally ignored the reference to the United States as a whole: (a) “No because Colorado and Arizona might not have children with guns,” (b) “No because it only talks about Chicago,” and (c) “I’d have to find out more about it, this story might not be true for all high schools.” Similar responses are given by a larger proportion of the Grade 9 students, again by many who had missed spotting the difficulty in their answers to the first part of the question. Some of these responses are stated in a more sophisticated manner: (a) “No because just because Chicago is like that doesn’t mean other places would be,” and (b) “No because the whole of the United States wasn’t surveyed, so we don’t know that the handgun situation is the same throughout the United States.”

It is also of interest to consider the responses made by students who continue to miss the main point of the questions. There are some who probably do not take the question seriously: (a) “I don’t want to be a teacher,” and (b) “No because I’d have a gun too.” Many accept and agree with the claims being valid outside the sample area. Some of these responses may be influenced by the students’ experiences of the United States on television: (a) “Yes because I wouldn’t want to take the risk of getting shot,” and (b) “Yes because Americans are insane.” Some also say they would teach in other areas but for reasons not associated with the sampling problem: (a) “No, because only 4% are taking them to school,” and (b) “No because it doesn’t say that these kids fired the gun, they were just showing off. Being tough.”

Analysis

In terms of defining levels of response appropriate for this question, it is the second part of the question which provides the foundation for building a picture of student understanding. Many students can provide a statistically acceptable answer to the second part but not the first. The reverse however does not occur; those who can answer the first part can always answer the second. Categories as given below seem appropriate for this item.

- No engagement with concepts:

  “I’d have a gun too.”
  “Rules in USA schools are terrible.”

- One peripheral statistical point:

  “They could be lying.”
  “Only 4% are taking guns to school.”

- Recognition with geographical cue (second part of question):

  “No because it only talks of Chicago.”
  “No because the whole of the United States wasn’t surveyed ...”

- Recognition without geographical cue (first part of question):

  “They are say 6 in 10 United States high school students when...”

Interpretation

It should be noted that this item is structured differently from most items used for assessment. Usually the first part of a question is easier, leading on to a more sophisticated second part, as in the previous pie chart item. The reason for structuring the sampling question in the way it was presented mirrors the manner in which the article might be used as a basis for classroom discussion. A teacher beginning a session on sampling might present the article and ask students to discuss the first question. It is likely that many responses, similar to those given here which were off the point, would come up during discussion. The teacher would then focus the class on one of the high level responses if it arose or move on to the second question as a means of leading students to an appreciation of the sampling problem. It is unlikely a teacher would ask the first question and if no adequate answers were forthcoming then ask, “What is the relationship of the Chicago poll to the claim about United States high school students?” Such a question would not allow for individuals to discover the relationship for themselves. The second question in this item might allow that to happen.

It is important that assessment items are designed to reflect the manner in which teaching and learning takes place. Hence it should not be considered impossible to ask a more difficult question first in some circumstances. This is especially so when the highest level of skills is being assessed.

SUMMATIVE ASSESSMENT WITH MEDIA ITEMS

If summative assessment is to be carried out in conjunction with the use of the items introduced above, then a marking scheme based on the levels in the “Analysis” section for the items can be developed. The increasing degree of sophistication can be associated with increasing integer values. This is a holistic scheme, which is a variation on that suggested by Charles et al. (1987). For the first part of the pie chart question assigning *0, 2, 4, 6*, to the four categories of response with the possibility of odd numbers for answers, displaying slightly less understanding or quality would be appropriate. If one were to score the entire item including both questions in a hierarchical fashion, it would appear reasonable to give the complete answer (including the “128.5” response for the second part an *8* and including the “61.5” response) a *7*. For the sampling item it would appear reasonable to assign quantitative scores of *0*, *2*, *4*, and *6* , respectively, for responses in this hierarchy.

The assigning of marks described here may seem relatively simplistic, but much effort has gone into the definition of levels reflecting increasing achievement. It is unlikely that summative assessment for students would consist only of numbers but also would include comments explaining why the numbers were assigned. This is the method employed by Garfield (1993) in reporting on project work in statistics.

IMPLICATIONS

It has been the purpose of this chapter to examine the levels of thinking required to interpret statistical information and claims made in social settings (and reported in the media), to relate these to the assessment of interpretive statistical thinking, to provide examples of how an assessment scheme operates, and to consider some of the classroom issues associated with such assessment. The goal is to achieve third-tier statistical literacy for all students in terms similar to those stated by Wallman (1993) in an address to the American Statistical Association.

> “Statistical Literacy” is the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions. (p.1)

There are several implications for the classroom teacher which arise from the need to assess interpretive statistical thinking and from the examples used in this chapter. Webb (1992, p. 667) claims that assessment which is integral to instruction embodies four features. Each of these is paraphrased in terms of the context of this chapter. First, the teacher must understand the structure of the statistical thinking hierarchy and use this structure to define expectations for learning. Second, the teacher must be sensitive to the processes students use to learn critical statistical thinking, the stages of development, and the processes available to facilitate this contextual statistical learning. Third, assessment is a process of gathering information about a student's knowledge about statistical thinking, about the structure and organization of the tiers of that knowledge, and about a student's cognitive processes, then giving meaning to the information obtained. This is the goal of the "Analysis" and "Interpretation" sections for the items introduced in this chapter, and it would be expected that teachers would acquire the ability to structure other assessment in a similar fashion using that model. Fourth, assessment employing items such as described in this chapter is used to make informed decisions about methods of instruction. Such instruction should be based on current information available about what a student knows and about what a student is striving to know. The aim is to assist each student in achieving higher levels.

Several other issues warrant mention with regard to the assessment of statistical thinking in the classroom. One relates to the number of students who are assessed in relation to a single outcome. Many curriculum documents (e.g., AEC, 1991; NCTM, 1989) suggest cooperative group work and report-writing in connection with objectives such as those discussed in this chapter. The assessment of such work, both within classrooms and on a larger scale, has not been addressed widely in the mathematics curriculum, let alone with respect to statistics. Assessment questions relate to the assignment of sub-tasks within a group, the degree of input provided by each participant, the quality of the final product produced by the group, and individual learning which has taken place as a result of the group activity. All of these questions will need to be addressed for items which begin with media extracts and involve cooperative work leading to assessment. Because much of the work in society which leads to media reports is the result of the efforts of teams of researchers, pollsters, and so forth, it is relevant to develop assessment techniques valid for teams as well as individuals in this context.

The combination of motivated students, well-informed teachers, relevant content and a useful scheme for assessment should ensure that the higher order thinking required for statistical literacy is achieved for most if not all students by the end of secondary school.