POST SECONDARY AND ADULT STATISTICAL LITERACY: ASSESSING BEYOND THE CLASSROOM

Jennifer J. Kaplan and Justin Thorpe
Department of Statistics and Probability, Michigan State University, United States of America
kaplan@stt.msu.edu

There is no question that an informed citizenry needs to be statistically literate. Many definitions for statistical literacy have been proposed. While there are certain similarities among them, consensus has not yet been reached. This paper defines adult statistical literacy as the set of skills and knowledge used by expert consumers of statistics and then provides a potential framework, based in the literature, to describe the components of statistical literacy. The paper then illustrates an assessment method, interpreting 40 responses to an authentic task through the lens of Watson and Callingham's Statistical Literacy Construct. Potential extensions are discussed.

INTRODUCTION

In her 1992 Presidential address to the annual meeting of the American Statistical Association (ASA), Katherine Wallman stated that citizens of our society do not have the necessary statistical understanding to evaluate the statistics information they encounter daily. She asserted that to address the problems associated with the lack of statistical understanding of the populace, trained statisticians should be deeply involved in the development of statistics and quantitative literacy programs for students of all ages. More recently statistics educators and statisticians have called for increased focus on the development of statistical literacy (Utts, 2003). Gal (2002) advises that more research is needed on students' and adults' literacy skills. In particular, he advocates that large educational systems and academic institutions provide the optimal forum for improving the level of statistical literacy achieved by students studying statistics. He asserts, however, that for improvements in statistical literacy to occur, the instructors of such courses must be given the proper resources and tools to help students to develop statistical literacy.

The definition of statistical literacy is still being refined (Rumsey, 2002) and many definitions for statistical literacy have been proposed. For example, Wallman (1993) states that statistical literacy is the ability to understand and critically evaluate statistical results that permeate daily life and the ability to appreciate the contribution statistical thinking can make in the decision making process. Gal (2003) defines statistical literacy as the ability to interpret, critically evaluate, and, when relevant, express opinions about statistical information, data-related arguments or stochastic phenomena. The college report of the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Project defines statistical literacy as understanding the basic language of statistics for example, knowing what statistical terms and symbols mean and being able to read statistical graphs, and understanding some fundamental ideas of statistics (GAISE, 2006). Carlson (2002), quoting report of the Programme for International Student Assessment (PISA) produced by the Organisation for Economic Co-operation and Development (OECD), asserts that, "*statistical* literacy is…the ability to put *statistical* knowledge and skills to functional use rather than just mastering them within a school curriculum (emphasis original)" (pg 4). Through a careful reading and synthesis of the literature, one finds that there is a convergence in the ideas that have been published that provides basic agreement about what a statistically literate person should know.

The view taken here is that taken by Gal (2003): statistical literacy is the set of statistical ideas that are used by "consumers, rather than producers, of statistical information" (p. 3). With that in mind, the aim of this paper is to provide a preliminary framework to describe statistical literacy, defined as the skill set needed to be an expert consumer of statistics. Second, the paper provides an example of how adult statistical literacy might be assessed outside the classroom, embedded within an authentic task, using the Statistical Literacy Construct (Watson & Callingham, 2003) to categorize responses by level.

In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Table 1. Preliminary Framework for Describing Statistical Literacy

Content Area	What People will Know and Do	Critical Questions (Gal, 2002)
Data, Data Collection and Experimental Design	Knowing why data are needed and how data can be produced (Gal, 2002 and MAA/ASA joint committee, 1992) • data are part of everyday life and an important aspect of the working world (Rumsey, 2002) • decisions made based on data can have a strong impact in our lives (Rumsey, 2002) • How to collect data (Rumsey, 2002 and GAISE, 2006) • Random *sampling* allows results of surveys and experiments to be extended to the population from which the sample was taken. (GAISE, 2006) • Random *assignment* in comparative experiments allows cause and effect conclusions to be drawn (Utts, 2003, GAISE, 2006) • Common sources of bias in surveys and experiments (Utts, 2003)	• Where did the data originate? What kind of study was it? Was the type of study appropriate? • Was a sample used? How was it sampled? How many subjects? Was there bias? • How reliable were the instruments or measure used?
Probability	Understanding basic notions of probability (Gal, 2002) • The idea that coincidences and very improbably events are not uncommon because there are so many possibilities (Utts, 2003) • Confusion of the inverse in conditional probability (Utts, 2003)	• What is the shape of the underlying distribution of the raw data? Does it matter? • How was this probabilistic statement derived?
Variability	Presence, Quantification and Explanation of Variability (MAA/ASA joint committee, 1992) • Variability is natural (Utts, 2003, GAISE, 2006) • Variability is predictable and quantifiable (GAISE, 2006)	• Are measures of variability reported? • Are small differences made to “loom large”? • Are there unusual values that distort the results?
Descriptive Statistics	Familiarity with basic terms and ideas related to descriptive statistics (Gal, 2002, Rumsey, 2002) • Knowledge of the basics and generating descriptive statistics (Rumsey, 2002) • Understanding that “normal” and “average” are not the same (Utts, 2003) • How to graph the data as a first step in analyzing data (GAISE, 2006) • How to interpret numerical summaries and graphical displays of data (GAISE, 2006)	• Are the reported statistics appropriate for this kind of data? • Is the graph drawn appropriately? • Are the descriptive statistics enough to answer the question of interest?
Conclusions and Inferences	Knowing how statistical conclusions or inferences are reached (Gal, 2002) • The difference between statistical significance and practical importance, especially for large sample sizes • The difference between finding “no effect” or “no difference” and finding no statistically significant effect or difference, especially for small samples.	• Are the claims made sensible and supported by the data? • Should additional information or procedures be made available to enable me to evaluate the sensibility of the argument? • Are there alternative interpretations that were not discussed?

WHAT IS STATISTICAL LITERACY? A PRELIMINARY FRAMEWORK

Through the lens of consumership and using five publications about statistical literacy or learning goals that include an outline or framework, a preliminary framework for describing statistical literacy has been developed. The five publications are Cobb (1992), GAISE (2006), Gal (2002), Rumsey (2002), and Utts (2003). Each of the five publications provides an outline or list of the elements of statistical literacy that the author deems important. In the case of the GAISE report, the authors provide a comprehensive list of the topics to be mastered by students in an undergraduate introduction to statistics course. Only the competencies that were related to consuming statistics were included in this framework. Gal (2002) provides in his outline of statistical literacy not only a list of the statistical knowledge that is necessary for statistical literacy, but also a sample of "worry questions" that embody the critical skills required of a statistically literate person. The worry questions presented represent the types of questions about statistical messages, particularly those that appear in the media that Gal expects a statistically literate person to ask.

Through open coding of the outlines presented in the publications, five content themes emerged: Data and Experimental Design, Probability, Variability, Descriptive Statistics, and Conclusions and Inferences. These are listed in the first column of Table 1. The specific content within each content theme, given in column two of Table 1, were taken directly from the outlines. In addition, the worry questions provided by Gal (2002) were incorporated as a second dimension for each of the 5 content areas and are given in column three of Table 1. The organization of the framework around content themes was used to create a description that would be understood and enacted in the classroom easily by statistics instructors. In this way the framework will provide the type of resource that Gal (2002) suggests in necessary for the improvement of teaching and learning in the area of statistical literacy.

ASSESSING ADULT STATISTICAL LITERACY

Watson and Callingham (2003) provide a hierarchical model to measure the statistical literacy construct in school-aged students that it is hypothesized can be overlaid on the framework given in Table 1 and used to describe levels of adult statistical literacy. This model is comprised of six levels: Idiosynratic, Informal, Inconsistent, Consistent/Non-critical, Critical and Critical Mathematical. The model specifies "task steps" exhibited by students at each level in the areas of sampling, terminology, chance, variation, average, graphs, inference and questioning claims. Sampling and terminology correspond roughly to the first row of Table 1: data, data collection and experimental design. Chance aligns with probability, variation with variability, average and graphs with descriptive statistics and inference and questioning claims with conclusions and inference. The levels are described in more detail in Table 2 using the task steps for sampling, terminology and inference as an example. The remainder of this section of the paper discusses the use of the Watson and Callingham model to describe responses of undergraduate students to a task designed to measure statistical literacy.

The task used to assess the consumption of statistics is the Email from Dad task described in Kaplan (2009). It is designed to assess the two content areas of data, data collection and experimental design and conclusions and inference. In brief, the subjects were asked to write a response to an email from their "father" asking about a discussion the father had with a doctor about a possible blood pressure medication for the subject's grandmother. The father's email describes the successful results of a clinical trial and includes negative anecdotal evidence from two of the father's acquaintances. A total of 40 responses to the task were included in the analysis, collected as part of two different protocols. Thirty of the responses were collected as part of a pencil and paper protocol that included two tasks from psychology, a reading comprehension task, and 13 multiple choice questions about statistical knowledge. The other 10 responses were collected at the beginning of a clinical interview after the same psychology tasks that appeared on the pencil and paper protocol. In addition to the written portion of the interview, subject role-played follow up conversations to the email from dad and discussed their reactions to two other statistics related articles. Only the written responses to the Email from Dad task were used in the analysis presented in this paper.

Table 2. Statistical Literacy Construct

Level	Description	Characteristic Task Steps	Descriptors of Adult Literacy Responses
Idiosyncratic	Idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables	unwilling to make predictions; domination of personal belief and experiences	Incorrect and possibly detrimental decisions; use of terminology with no indication of understanding
Informal	colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, graph and chance calculations.	use of story telling or pattern recognition	Use of Story Telling; trusting of the doctor or the FDA
Inconsistent	selective engagement with context, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas	use of data driven stories without justifications; focus on peripheral rather than salient features	Addresses statistical terms, discuss experimental design and mention p-value with at least one correct
Consistent/ Non-critical	appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in change settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics	engaging with context, questioning data collection; partial recognition of salient features	Discussion of design is dealt with professionally, discussion of p-value is closer to correct, recognizes justifications
Critical	critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation or chance and appreciation of variation	giving appropriate decisions and statistical justifications, focus on central issues	Clearly defines terms, discusses affordances and limitations, defends the p-value.
Critical Mathematical	critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need to uncertainty in making predictions, and interpreting subtle aspects of language.	acknowledging the probabilistic subtleties in statistical conclusions; asking salient questions about relationships	Not seen in the data due to lack of evidence that subjects correctly understood mathematical underpinnings of inference.

All of the subjects had taken a non-calculus based one-semester introductory statistics course offered by the department of mathematics at a large research institution in the U.S. Southwest. Thirty-two of the subjects (75%) were women. Twenty-four (60%) reported having a G.P.A. of 3.5 or higher. Thirty-three subjects (82.5%) reported earning an A or B in statistics. The four concentrations with the highest number of participants were pre-pharmacy, 10 (25%), nursing, 8 (20%), advertising, 6 (15%) and psychology, 4 (10%). The number of words used in the responses to the Email from Dad task was roughly normally distributed with a mean of 116 words and standard deviation 56 words.

The data were categorized by two independent raters. Prior to coding, the raters read Watson & Callingham (2003) and created a categorization scheme based on the general description of the six levels of the Statistical Literacy Construct and the characteristic task steps reported in the publication for the two content areas associated with the task: data collection and inference. These descriptions can be found in columns 2 and 3 of Table 2. In the first round of coding, the raters only agreed on 11 (27.5%) of the responses. The codes for 18 (45%) responses differed by only one level. The raters then discussed each response individually until consensus was reached on every response. From this discussion, descriptions of the six levels of statistical literacy, as applied to the Email from Dad task, were developed. These descriptions can be found in the final column of Table 2.

The seven responses at Level 1–Idiosyncratic are characterized by incorrect responses that could possibly lead to the making of detrimental decisions and indicate no understanding of the statistical terminology. Some examples are telling the father to take Grandma off the medication because the p-value less than .001 means that the chance that the medication will lower Grandma's blood pressure is less than 1% or that the double blind design means that the researchers did not know who was getting which medication and, in fact, they might not have been testing the medication at all. The nine responses at Level 2–Informal are characterized by either the use of story telling without reference to statistical information or an indication of trust in an authority. Responses at this level include advice to trust the doctor or consult with him again about side effects based on the anecdotal evidence provided. In other responses at this level, subjects state that the results of the test indicate that the medication is effective, but there is no attempt to explain how or why that is the case.

In Level 3–Inconsistent, the decisions are supported by statistical data. In general, however, the sixteen subjects in this category do not demonstrate a clear understanding of the statistics that are mentioned. Subjects at this level tend to mention the p-value, although the interpretation is incorrect, give a decision on the sample size being large enough, and correctly explain the meaning of double blind and/or randomization. These responses have the feel of students moving through an inference checklist they learned in class, but showing little understanding beyond the learned checklist. In Level 4–Consistent Non-critical, the three subjects discuss experimental design issues more professionally than at the previous level. These responses still suffer from a lack of connection between the interpretations of the p-value and the hypotheses. Instead, the subjects discuss in a vague manner either statistical significance or whether the results could be due to chance. These errors in p-value interpretation, however, would cause fewer issues in general decision making than those that appear at previous levels. That said, responses at this level still provide little evidence that the subject could respond to a similar task that had a different context.

The three responses at Level 5–Critical link p-values to hypotheses and/or probability in a specific way and are coherent in their discussion of experimental design. At this level, discussions of the anecdotes are tied to discussions of variability and/or experimental design. The data did not contain any responses that were categorized as Level 6–Critical Mathematical. This was largely because the raters could not be certain, from the writing samples, that any of the subjects understood the mathematics behind inference and p-values correctly and in a way that would allow such knowledge to be applied in alternate contexts. In addition to these levels, there were two subjects who responded to the email by admitting that they had not retained enough information from their statistics class to feel competent to provide their father with useful information. These were not categorized as Level 1–Idiosyncratic because of the subjects' realization of their lack of competence. At level 1, subjects tended to give opinions without recognizing how little of the situation they actually understood.

CONCLUSIONS

There are two directions suggested by this pilot study for future research in this area. The first is the validation of the framework to describe Adult Statistical Literacy and the second is an expert-novice study designed to saturate a model of the Statistical Literacy Construct as it applies to Adult Statistical Literacy. This paper provides an opportunity to begin the process of validating the framework to describe Adult Statistical Literacy, as will planned interviews with statisticians and graduate students in statistics. The results from the coding of the pilot data on the Email from Dad task indicate that the Statistical Literacy Construct developed based on school-aged children can be extended to assess Adult Statistical Literacy. The lack of initial consistency by the independent raters indicates that the task of applying the Watson and Callingham model to open-ended tasks designed to assess Adult Statistical Literacy is not trivial in nature. The ability of the two raters to come to agreement and describe the six levels of Statistical Literacy based on the data, however, indicates that the Construct has potential as a basis for assessing, categorizing and describing Adult Statistical Literacy.

Once the framework to describe Adult Statistical Literacy has been validated within the statistics and statistics education communities, more tasks to assess Adult Statistical Literacy will be developed. Using these tasks we will conduct a cross sectional study, using subjects across the continuum from novice to expert statistician in order to either validate the Watson and Callingham model or to refine the model as necessary to reflect differences between youth and adult statistical literacy. At the time of writing the research team has already planned to collect cross sectional data from undergraduate statistics majors and graduate students in statistics, using the Email from Dad task, and then to code the data using the model discussed in this paper. These results will be used to saturate and refine the model presented to reflect a range of statistical literacy abilities. The overarching goal of the research program suggested by this pilot work is to provide a description of expert adult statistical literacy to use as a target for instruction which, following the suggestion of Wallman (1993) utilizes the expertise of statisticians as a basis and that of Gal (2002) can be implemented via large educational systems and academic institutions.

REFERENCES.

Carlson, B. (2002). Preparing workers for the 21st Century: The importance of statistical compentencies. In B. Phillips (Ed.) *Proceedings of the 6th International Conference on Teaching Statistics*. Online: www.stat.auckland.ac.nz/~iase/publications.php?show=1.

Cobb, G. (1992). Teaching Statistics. In L. Steen (Ed.), *Heeding the Call for Change: Suggestions for Curricular Action* (pp. 3 – 43). Washington, D.C: Mathematical Association of America.

GAISE (2006). College Report of the Guidelines for Assessment and Instruction in Statistics Education Project. Online: *www.amstat.org/education/gaise/GAISECollege.htm*.

Gal, I. (2003). Expanding the conceptions of statistical literacy: An analysis of products from statistics agencies. *Statistics Education Research Journal*, *2*(1), 3 – 21. *http://fehps.une.edu.au/serj*

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1 – 51.

Kaplan, J. J. (2009). Effect of Belief Bias on Undergraduate Students' Reasoning about Inference. *Journal of Statistics Education, 17(1).* Online: www.amstat.org/publications/jse/v17n1/kaplan.html.

Rumsey, D. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). Online: www.amstat.org/publications/jse/v10n3/rumsey2.html

Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, *57*(2), 74 – 79.

Wallman, K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, *88*(421), 1 – 8.

Watson, J. & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3 – 46. Online://fehps.une.edu.au/serj.

Wild, C. J. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223 – 265.