Undergraduate Students' Statistical Literacy in Media Contexts

Samuel Waters
University of Northern Colorado

As the uses of data in society continue to expand, it is vital to prepare students to engage with data in meaningful and critical ways. This study used a course-wide assessment in an introductory statistics course to explore undergraduate students' statistical literacy in media contexts based on Watson and Callingham's (2003) six-level construct. Statistical analysis found student performance on non-media tasks had limited association with performance on media-based tasks. Students demonstrated higher levels of statistical literacy when provided with specific prompts. Findings suggest the need to develop students' propensity for critical questioning when engaging with statistics outside classroom settings through intentional incorporation and discussion of media items in introductory statistics courses.

Keywords: Statistics, Introductory Courses

For most adults, the news media is the primary source of information about social and scientific issues (Bissonnette et al., 2021). Statistical and mathematical products in the news take on many forms, requiring multiple inter-related skills to appropriately comprehend and evaluate (Gal & Geiger, 2022). As the pervasiveness of data in everyday life continues to increase, statistics educators are exploring how to prepare students to engage with data in meaningful and critical ways (Burrill & Pfannkuch, 2023). Undergraduate introductory statistics courses provide a natural place where the development of these skills could occur. Driven by this need, the purpose of this quantitative study was to understand the statistical literacy of students in an introductory statistics course within media contexts. Statistical literacy is defined as a multi-leveled construct building from idiosyncratic task steps (Level 1) where no mathematical or statistical thinking is occurring, to critical mathematical task steps (Level 6) which involve critical engagement and questioning with the statistics in context (Watson & Callingham, 2003).

Background

While early statistics education research focused on resolving persistent challenges in developing reasoning about specific statistical ideas (Garfield & Ben-Zvi, 2007), current trends looking to the future of statistics education research include broader issues of content, pedagogy, and ethics (Burrill & Pfannkuch, 2023). One of the four emerging trends Burrill and Pfannkuch identified is social statistics, which encompasses the importance of attending to risk, critical literacy, and communication in contexts such as the media and advocacy. Burrill and Pfannkuch noted that while statistical literacy from the perspective of societal needs was recognized by Gal (2002) as well as Watson and Callingham (2003), the ideas gained limited attention. Existing research on statistical literacy spans a variety of methods and foci. Previous quantitative studies typically used multiple choice tests to evaluate the statistical literacy of students in introductory statistics courses (e.g., Chance et al., 2022). Such studies frequently show students struggle to demonstrate mastery of topics such as sampling, variation, and study design even after completing a statistics course (delMas et al., 2007; McLauchlan & Schonlau, 2016).

Many statistics education researchers intentionally investigate statistical literacy in media contexts. Watson (1997) advocated for the use of media-based tasks in courses to assess statistical thinking in social settings outside the classroom. Other studies utilized media-based tasks to investigate student statistical literacy in general (e.g., Kaplan & Thorpe, 2010) or understanding of specific topics such as sampling (Watson & Moritz, 2000). Budgett and Pfannkuch (2010) found incorporating media into classroom activities and assessments to be effective for college students in non-quantitative majors.

Simulation-based curricula (SBC) have been growing in popularity in introductory statistics courses (Tintle et al., 2015). SBC provide an alternative to using theoretical distributions to calculate confidence intervals and p-values by simulating repeated samples to generate bootstrap and randomization distributions (Lock et al., 2021). Advocates of SBC assert they allow students to see the logic and power of statistical inference and find a balance between statistics as proof and statistics as magic (Tintle et al., 2015). Researchers identified improvements in students' statistical literacy when they engage with SBC compared to engaging in more theory-based curricula (e.g. Chance et al, 2022; Maurer & Lock, 2016). Other studies noted how simulation tools contributed to student understanding of sampling (Findley & Lyford, 2019), p-values (Reaburn, 2014), and inference (Noll et al., 2023).

While current research calls for increased use of simulation-based methods in introductory statistics courses (Tintle et al., 2015), research is less clear about how SBC help prepare students to engage with statistics in the increasingly complex media environment. This study brings together research on statistical literacy, SBC, and statistics in news media to understand how students in introductory statistics courses utilizing SBC are prepared to engage with statistical information in media settings. Specifically, this study addresses the research questions:

1. At what level of statistical literacy do students in an introductory statistics course apply statistical topics when critiquing statistics in the news media?
2. What relationship exists between students' scores on non-media-based tasks and their levels of statistical literacy on media-based tasks?

Conceptual Framework

This study adopted Watson and Callingham's (2003) six-level construct of statistical literacy, grounding statistical literacy in the need for understanding, interpreting, and communicating in real-word contexts such as the news media. Individual levels are characterized by the types of task steps students take while engaging with statistical tasks. Progressing to higher levels of the construct statistical constructs requires more nuanced statistical thinking, clearer justifications or reasoning, and deeper engagement with the task's context. Level 1 and Level 2 involve mostly non-statistical thinking, instead relying on personal experiences and intuition. Level 3 and Level 4 utilize statistical terminology and arguments with increasing amounts of justification and connections to the task's context. Level 5 and Level 6 incorporate critical engagement and quantitative interpretations concerning the data and conclusions.

This study's conceptual framework draws on Gal and Geiger's (2022) categories of statistical and mathematical products (StaMPs). Gal and Geiger conducted a content analysis of over 300 media items and identified nine categories that encompass entities appearing in media items. This study focuses on five of the nine categories to investigate students' statistical literacy of statistical topics supported by SBC: (a) descriptive quantitative information; (b) models, predications, causality, and risk; (c) data quality and strength of evidence; (d) demographics and comparative thinking; and (e) critical demands. The nine categories provide a basis for what students should be prepared for when engaging with statistics in news media contexts.

Methods

This study took place at a mid-sized, regional university in the Rocky Mountain Region of the United States. The university is an emerging Hispanic Serving Institution with a student population that is 62% White and 24% Latine/Hispanic with 42% of incoming students being first generation college students. Introductory statistics courses utilize a textbook and materials which address simulation-based methods for calculating confidence intervals and p-values (Lock et al., 2021). All 189 students enrolled in seven sections of introductory statistics were asked to complete an assessment as part of the course for a completion grade. Overall, 123 students (65.1%) completed at least one part of the assessment and consented for their responses being included in the study.

Each participant completed a three-part assessment using Qualtrics during the 12th week of the semester. Part 1 consisted of six multiple-choice questions taken from the Basic Literacy in Statistics (BLIS) assessment (Ziegler & Garfield, 2018). The BLIS assessment was selected because it was developed for introductory statistics courses at the postsecondary level that include SBC. Selected items were not embedded in media contexts and assessed participants' understanding of three statistical topics that are encountered in the news: sampling, variation, and confidence intervals. Variation and confidence intervals were selected as they are topics where SBC have been shown to be particularly effective (Chance et al., 2022). Parts 2 and 3 were media-based tasks asking participants to read a headline and short article containing information related to the same three statistical topics from Part 1. Part 2's article discussed results from a recent drug trial and asked participants to evaluate the presented argument but gave no additional prompts about what to talk about (i.e., unprompted). Part 3's article reported results of a poll and asked participants to respond to prompts referring to specific features and statistics within the article (i.e., prompted).

Participant responses were scored to facilitate statistical analysis. Multiple choice items were scored on correctness and given a zero or one, forming six variables (MCQ1, MCQ2, etc.). Topic scores (i.e., MC_{Samp}, MC_{Var}, and MC_{CI}) were calculated based on the number of correct responses on each topic (out of two) and an overall multiple-choice score (i.e., $MC_{Overall}$) indicated the total number of correct responses (out of six). Scores for the unprompted (UP) and prompted (P) media tasks were determined using Watson and Callingham's (2003) six-level construct of statistical literacy. For each task, participant responses were assigned a level score from 0 to 6 for each of the three statistical topics based on the statistical literacy level their response addressed (i.e., UP_{Samp}, UP_{Var}, UP_{CI}, P_{Samp}, P_{Var}, and P_{CI}). A score of 0 was assigned to responses that did not address the statistical topic at all, either because it was not mentioned for the unprompted task or left blank for the prompted task. Based on the three topic scores, each participant's response was assigned an overall level of statistical literacy (i.e., $UP_{Overall}$ and $P_{Overall}$). In total, there were 18 variables: (a) six individual multiple-choice question scores; (b) three topic scores for sampling, variation, and confidence intervals on each part (nine total); and (c) three overall statistical literacy scores (one for each part).

Response scores were imported into R for statistical analysis. Summary statistics were calculated for each variable. Since scores were not normally distributed, variables were treated as ordinal when conducting statistical inference. Non-parametric tests such as the Wilcoxon Signed-Rank test and Kruskal-Wallis Rank test were performed to compare scores. Correlational analysis was performed to see how different variables related to each other. Spearman's rho correlations were used due to the non-continuous nature of the variables (Agresti, 2007). Additionally, analysis was performed to determine if multiple-choice questions were strong predictors of statistical literacy level on media tasks. Ordinal logistic regressions were used to incorporate the ranked property of the variables and explore how multiple-choice responses predict the odds of higher statistical literacy (Osborne, 2017).

Results

Performance on the multiple-choice questions ranged from 69.1% correctly answering MCQ1, which dealt with identifying a sample and population, to 16.3% correctly answering MCQ6, which asked participants to interpret the center of a confidence interval. Most students answered one to three questions correctly with the mean $MC_{Overall}$ score being 2.285 correct out of six. No participant answered all six questions correctly. Participants had the most correct responses on sampling questions, followed by variance and, lastly, confidence intervals.

Table 1provides a summary of participant scores on the unprompted and prompted media tasks. For both tasks, participants demonstrated higher mean statistical literacy scores related to sampling than for variation or confidence intervals. The highest mean was P_{Samp} ($M = 3.546$) and the lowest mean was UP_{CI} ($M = 1.822$). Overall, participants demonstrated lower levels of statistical literacy on the unprompted task compared to the prompted task on all three statistical topics and the overall score. Wilcoxon signed-rank tests were performed to assess whether the medians of the distributions of scores on the unprompted and prompted tasks were equal. Results of the tests indicated significantly greater scores on the prompted tasks for sampling ($p < .001$), variation ($p < .001$), confidence intervals ($p = .002$), and the overall score ($p < .001$). Participants also had more zero scores on the unprompted task compared to the prompted task, indicating participants were not addressing certain pieces of information within the unprompted task. For example, 40 participants received a zero on UP_{CI} as their response did not include the comparison of values discussed in the article.

Table 1: Summary of Statistical Literacy Scores on Media Tasks

Task	Sampling		Variation		CI		Overall	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Unprompted [a]	2.237	1.616	2.051	1.524	1.822	1.583	2.839	0.961
Prompted [b]	3.546	1.314	2.944	1.478	2.537	1.632	3.389	1.134

[a] n = 118. [b] n = 108.

To investigate the relationship between participant scores across topics and task type, Spearman rho correlations were calculated for all scores (see Table 2). Analysis focused on the correlations between non-media task scores (i.e., multiple choice) and media task scores to see what relationship existed. All correlations were weak, with the strongest correlation being $\rho = .260$ between $MC_{Overall}$ and UP_{CI}. All correlations were positive, indicating greater scores in one area were associated with greater scores in other areas. Eight of the 32 correlations were significant. MC_{Var} was not significantly associated with any of the media-tasks scores. $MC_{Overall}$ had many of the strongest correlations, including significant correlations with the $UP_{Overall}$ ($\rho = .216$) and $P_{Overall}$ ($\rho = .226$). The only multiple-choice topic score significantly correlated with a corresponding media-task topic score was for MC_{CI} and UP_{CI} ($\rho = .193$).

Due to potential similarities between the prompts in a multiple-choice setting and the prompted media task, additional analysis focused on the relationships between those sets of questions. For example, MCQ1 asked students to identify the sample and population of a provided context. Similarly, the first prompt of the prompted media task asked students to identify the sample and population. Identifying both correctly would correspond to a statistical literacy score of at least Level 3 for P_{Samp}. While participants who answered MCQ1 correctly had a higher mean P_{Samp} score (M = 3.688 vs. M = 3.143), a Kruskal-Wallis rank-sum test found no significant difference in the medians of the two groups (p = .058).

Table 2: Spearman Correlation Matrix of Multiple-Choice Scores and Media Task Scores

		Unprompted				Prompted			
		Samp	Var	CI	Overall	Samp	Var	CI	Overall
Multiple Choice	Samp	.156	.144	.157	.192*	.145	.065	.218*	.234*
	Var	.139	.092	.147	.124	.113	.071	.043	.063
	CI	.179	.003	.193*	.073	.087	.068	.083	.157
	Overall	.248**	.127	.260**	.216*	.186	.118	.169	.226*

$^*p < .05$. $^{**}p < .01$.

The significant correlation between $MC_{Overall}$ and $P_{Overall}$ was explored further to understand the relationship between students' scores on non-media versus media-based tasks. A Kruskal-Wallis rank-sum test was performed to determine if there was a difference in the median $P_{Overall}$ score between groups based on $MC_{Overall}$ score. The test reported no significant difference (p = .131), suggesting scoring greater on the multiple-choice section did not correspond to greater distributions of statistical literacy scores on the prompted media task. An ordinal logistic regression model was fit to the two scores which similarly found different $MC_{Overall}$ scores were not significant predictors of $P_{Overall}$ scores. These results are reflected in participant scores as there were 21 participants who had lower $MC_{Overall}$ scores (score of two or less) but demonstrated greater levels of statistical literacy in the prompted media task (Level 4 or greater). Similarly, six participants had higher $MC_{Overall}$ scores (score of three or more correct) but did not demonstrate higher levels of statistical literacy in the prompted media task (Level 2 or lower).

Although $MC_{Overall}$ was not found to be a good predictor for the prompted media task, another ordinal logistic regression was performed to assess the predictive power of the individual multiple-choice questions. A binary variable for each question (correct or incorrect) was used in a model to predict overall prompted score. Table 3 includes the fitted coefficients for the regression. MCQ1 (sample and population) and MCQ5 (confidence interval estimation) were significant predictors of overall prompted score (p = .027 and p = .048 respectively). The odds ratio for MCQ1 indicates people who answer MCQ1 correctly were 2.569 times more likely to be in the next higher level of statistical literacy in the prompted task (95% CI [1.118-6.012]). Similarly, people who answer MCQ5 correctly were 2.053 times more likely to be in the next higher level of statistical literacy in the prompted task (95% CI [1.013-5.223]).

Table 3: Ordinal Logistic Regression for Multiple Choice Questions Predicting Prompted Task Scores

						95% Confidence Interval	
Question	Estimate (β)	*SE*	*t*-score	*p*-value	Odds Ratio	Lower	Upper
MCQ1	0.943	0.428	2.205	.027	2.569	1.118	6.012
MCQ2	0.603	0.458	1.317	.188	1.828	0.745	4.515
MCQ3	0.295	0.389	0.759	.448	1.343	0.627	2.893
MCQ4	-0.297	0.379	-0.782	.434	0.743	0.352	1.562
MCQ5	0.719	0.363	1.979	.048	2.053	1.013	4.223
MCQ6	-0.439	0.536	-0.820	.412	0.645	0.222	1.840

Discussion and Conclusion

The study found participants in introductory statistics courses applied statistical topics at varying levels of statistical literacy when engaging with the news media. Participants had difficulties correctly answering multiple-choice questions from a standard assessment (i.e., BLIS). The mean overall multiple-choice score was 2.285 out of six and only 16.3% selected the correct interpretation of a provided confidence interval. These results are consistent with ongoing challenges in introductory statistics courses McLauchlan and Schonlau (2016) identified where students do not always demonstrate improved statistical literacy after completing a course. While Chance et al. (2022) found SBC improved student outcomes related to topics such as variation and confidence intervals, this study shows challenges persist. Participants performed better on the sampling topic questions with more than two-thirds correctly identifying a sample and population from a provided context.

Results from the media-based tasks indicated that participants typically demonstrated significantly higher levels of statistical literacy when receiving specific prompts compared to an unprompted, open-ended question. Mean scores for the prompted task were mostly at Level 3 or higher, indicating students were more often engaging in statistical thinking and justification. However, mean scores for the unprompted task were mostly around Level 2, which corresponds to more intuitive thinking and/or incompleteness in statistical engagement. The results indicate a gap that needs to be addressed. Students have knowledge and abilities they are capable of applying but are not always doing so at the same level on their own (unprompted). For example, without prompts, several students (about one-third) did not regard key statistical information such as the comparison of the effectiveness of the two drugs being discussed in the article. In most real-life situations, such as engaging with the news media, information does not come with prompting questions. The results emphasize the need to facilitate students' development of what Gal (2002) described as *worry questions* to raise when assessing the quality of information.

Results of correlational analysis on the different task scores found only weak, positive correlations between scores on non-media multiple-choice questions and the level of statistical literacy on media-based tasks. The results indicate that demonstrating statistical knowledge in one type of task was not associated with demonstrating the same knowledge in other types of tasks. One specific example was that there was no significant difference in the distribution of sampling scores on the prompted task based on whether a participant correctly identified the sample and population in MCQ1. More broadly, about 30 students scored notably higher on one type of task compared to the other. Watson (1997) noted difficulty in students demonstrating greater levels of statistical literacy in multiple-choice settings. This study's findings indicate that multiple-choice questions may provide a limited picture of students' knowledge and abilities.

Ordinal logistic regression indicated correctly answering MCQ1 (identifying a sample and population) or MCQ5 (identifying what a confidence interval was estimating) were significant predictors in demonstrating higher overall levels of statistical literacy in the prompted media task. MCQ1 being the strongest predictor could indicate the importance of the concept. Understanding the difference between a sample and a population is a critical component of statistical inference and may contribute to students' ability to demonstrate greater levels of statistical literacy with the variation and confidence interval aspects of the task.

Limitations

The study has limitations to consider when interpreting these results. The assessment took place during Week 12 of the semester. While students had been assessed on the relevant content, students may have improved their statistical literacy related to the topics in this study by the end of the semester. Furthermore, the study only assessed students' understanding of sampling, variation, and confidence intervals which may not represent the entirety of their statistical literacy. The assessment contributed to participants' course grade, but only as a completion grade. Therefore, not all students may have responded at a level representative of their highest ability. There are possible limitations related to decisions about scoring such as the inclusion of the zero score. Importantly, zeros indicated the topic ideas students did not engage with for the tasks; zeros did not indicate participants had no statistical literacy related to the topic.

Implications

The study emphasizes the importance of supporting students' applications of statistics in real-life settings and the practice of critical questioning. Instructors and course developers should consider how to incorporate media tasks into class activities to provide opportunities for students to practice working in these contexts. Such tasks allow for discussions surrounding how statistics are presented in the news media and how the presentation can differ from textbooks and academic journals. These practices can help prepare students to engage with statistics in social settings, such as the news media (Burrill & Pfannkuch, 2023). When assessing students' statistical literacy, instructors should consider multiple forms of assessment and contexts to provide students with different opportunities to demonstrate what they know.

Further research should continue to explore how students approach statistics in media contexts. Qualitative research such as task-based interviews can provide additional insights into students' thinking and approaches to media items. In particular, research can investigate what students pay attention to or find important in these settings and explore the kinds of resources students draw on when evaluating statistical claims. Teaching experiments or other methods could be implemented to see what kinds of tasks, prompts, and activities are most effective in promoting critical questioning and higher levels of statistical literacy.

References

Agresti, A. (2007). *An introduction to categorical analysis* (2nd ed.). John Wiley & Sons, Inc. http://doi.org/10.1002/0470114754

Bissonnette, M., Chastenay, P., & Francoeur, C. (2021). Exploring adolescents' critical thinking aptitudes when reading about science in the news. Journal of Media Literacy Education, 13(1), 1-13. https://doi.org/10.23860/JMLE-2021-13-1-1

Budgett, S., & Pfannkuch, M. (2010). Using media reports to promote statistical literacy for non-quantitative majors. In C. Reading (Ed.), Data and context in statistics education: towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics. Ljubjana, Slovenia. https://iase-web.org/documents/papers/icots8/ICOTS8_7G1_BUDGETT.pdf?1402524971

Burrill, G., & Pfannkuch, M. (2023). Emerging trends in statistics education. ZDM – Mathematics Education. https://doi.org/10.1007/s11858-023-01501-7

Chance, B., Tintle, N., Reynolds, S., Patel, A., Chan, K., & Leader, S. (2022). Student performance in curricula centered on simulation-based inference. Statistics Education Research Journal, 21(3), 1-24. https://doi.org/10.52041/serj.v21i3.6

delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. Statistics Education Research Journal, 6(2), 28-58. https://doi.org/10.52041/serj.v6i2.483

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. International Statistical Review, 70(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x

Gal, I., & Geiger, V. (2022). Welcome to the era of vague news: A study of the demands of statistical and mathematical products in the COVID-19 pandemic media. Educational Studies in Mathematics, 111(1), 5-28. https://doi.org/10.1007/s10649-022-10151-7

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. International Statistical Review, 75(3), 372-396. https://doi.org/10.1111/j.1751-5823.2007.00029.x

Findley, K., & Lyford, A. (2019). Investigating students' reasoning about sampling distributions through a resource perspective. Statistics Education Research Journal, 18(1), 26-45. https://doi.org/10.52041/serj.v18i1.148

Kaplan, J. J., & Thorpe, J. (2010). Post secondary and adult statistical literacy: Assessing beyond the classroom. In C. Reading (Ed.), Data and context in statistics education: towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics. Ljubjana, Slovenia. https://iase-web.org/documents/papers/icots8/ICOTS8_5E3_KAPLAN.pdf

Lock, R., Frazer Lock, P., Lock Morgan, K., Lock, E., & Lock, D. (2021). Statistics: Unlocking the power of data (3rd ed.). John Wiley & Sons.

Mauer, K. & Lock, D. (2016). Comparison of learning outcomes for simulation-based and traditional inference curricula in a designed educational experiment. Technology Innovations in Statistics Education, 9(1). https://doi.org/10.5070/T591026161

McLauchlan, C., & Schonlau, M. (2016). Statistical literacy in the classroom: Should introductory statistics courses rethink their goals? Statistics, Politics, and Policy, 7(1-2), 99-115. https://doi.org/10.1515/spp-2017-0001

Noll, J., Kirin, D., Clement, K., & Dolor, J. (2023). Revealing students' stories as they construct and use a statistical model in TinkerPlots to conduct a randomization test for comparing two groups. Mathematical Thinking and Learning, 25(1), 44-63. https://doi.org/10.1080/10986065.2021.1922858

Osborne, J. W. (2017). Simple linear models with polytomous categorical dependent variables: Multinomial and ordinal logistic regression. In, *Regression and linear modeling: Best practices and modern methods* (pp. 133-156). Sage Publications, Inc. https://doi.org/10.4135/9781071802724

Reaburn, R. (2014). Introductory statistics course tertiary students' understanding of p-values. Statistics Education Research Journal, 13(1), 53-65. https://doi.org/10.52041/serj.v13i1.298

Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., & VanderStoep, J. (2015). Combating anti-statistical thinking using simulation-based methods throughout the undergraduate curriculum. The American Statistician, 69(4), 362-370. https://doi.org/10.1080/00031305.2015.1081619

Watson, J. (1997). Assessing statistical thinking using the media. In Gal, I. & Garfield, J. B. (Eds.), The assessment challenge in statistics education (pp. 107-121). IOS Press.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. Statistics Education Research Journal, 2(2), 3-46. https://doi.org/10.52041/serj.v2i2.553

Watson, J., & Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. The Journal of Mathematical Behavior, 19(1), 109-136. https://doi.org/10.1016/S0732-3123(00)00039-0

Ziegler, L., & Garfield, J. (2018). Developing a statistical literacy assessment for the modern introductory statistics course. Statistics Education Research Journal, 17(2), 161-178. https://doi.org/10.52041/serj.v17i2.164