# COMPUTES: DEVELOPMENT OF AN INSTRUMENT TO MEASURE INTRODUCTORY STATISTICS INSTRUCTORS' EMPHASIS ON COMPUTATIONAL PRACTICES

CHELSEY LEGACY
*University of Minnesota*
*legac006@umn.edu*

ANDREW ZIEFFLER
*University of Minnesota*
*zief0002@umn.edu*

ELIZABETH BRONDOS FRY
*St. Catherine University*
*ebfry704@stkate.edu*

LAURA LE
*University of Minnesota*
*free0312@umn.edu*

**ABSTRACT**

*The influx of data and the advances in computing have led to calls to update the introductory statistics curriculum to meet the needs of the contemporary workforce. To this end, the COMputational Practices in Undergraduate TEaching of Statistics (COMPUTES) instrument was developed to measure the extent to which computation practices—specifically data, simulation, and coding practices—are included in the introductory statistics curriculum. Data from 236 instructors were used in a psychometric analysis to evaluate the latent structure underlying instructors' response patterns and understand the quality of the instrument items. Responses were also examined to determine whether computational practices are being emphasized differently across institutional settings. Results suggest the latent structure was best captured using a correlated multidimensional model and most items were contributing information to the measurement process. Across institutional settings, curricular emphasis related to data and simulation practices seem quite similar, while emphasis on coding practices differs.*

***Keywords:*** *Statistics education research, computational thinking, assessment, instructional practices*

## 1. INTRODUCTION

Paralleling the metamorphosis of Ronald Miller in *Can't Buy Me Love* (Rash, 1987), statistics has gone from totally geek, to totally chic. Companies such as Google and Facebook have not only made a degree in statistics lucrative, but also made it popular (e.g., Hardy, 2012). Those in technology fields are not alone in their awareness of the power of statistics to understand the growing availability of data. Many non-tech-based companies and academic and non-profit institutions are also looking to hire employees with data and statistical acumen (Business Higher Education Forum and PricewaterhouseCoopers, 2017). As more employers come to see the ability to work with data and computational literacy as fundamental for those entering the workforce, colleges and universities will no doubt encounter more students interested in acquiring these skills (National Academies of Sciences, Engineering, and Medicine, 2018).

A decade ago, Nolan and Temple Lang (2010) called for reforming the statistics curriculum to focus on computation and data practices. Since that time, others have echoed this sentiment (e.g., Gould, 2017; Horton & Hardin, 2015; Kaplan, 2018) and the American Statistical Association (2017) has

*Statistics Education Research Journal, 21(1), Article 7. https://doi.org/10.52041/serj.v21i1.63*

directly addressed data and computation practices as fundamental skills within statistics curricula. There seems to be a growing consensus that computing and data practices are core skills for all students enrolled in a statistics course. Outside of the United States, educators and researchers have also been promoting data and computation practices in the curriculum (e.g., Engel, 2017; Engel et al., 2019; Ridgway, 2016, Ridgway et al., 2018). These practices have been included in several international curricular efforts and frameworks. For example, the Pro Civic Stat project in Europe has focused on promoting statistical literacy by using real world social contexts that are often multivariate (Gal et al., 2016; Schiller & Engel, 2016). Similarly, educators in New Zealand have integrated computation and authentic multivariate data and visualization within curriculum materials at all levels of education using thoughtfully designed software such as INZight and VIT (Forbes, 2014; Forbes et al., 2014). Other global efforts include the International Data Science in Schools Project (IDSSP Curriculum Team, 2019), a project to support and promote the teaching of data science in schools. This project has produced a set of curricular frameworks for teaching data science at the secondary level. Though efforts to implement more data focused statistics practices are evident in other countries, it is unclear whether instructors in the United States have incorporated these ideas into introductory statistics courses (Cobb, 2015; Horton & Hardin, 2015).

As computation and data practices become more prevalent in the contemporary workforce, it is important to be able to measure the degree to which instructors are introducing these ideas in the classroom. This is especially true given computation's role as a gateway to STEM fields (Holdren & Lander, 2012) and to more lucrative employment opportunities, especially for women and other underrepresented groups (Melguizo & Wolniak, 2012; U.S. Department of Commerce, 2017). Given the large numbers of students enrolling in introductory statistics courses—nearly three-quarters-of-a-million in the fall of 2015 (Blair et al., 2018)—including computation in these courses may help prevent disparities in students' computational learning opportunities. To that end, the goal of this research is to present the development of an instrument that can be used to measure the extent to which computation and data practices are included in the introductory statistics curriculum.

The influx of data and advances in computing have led to calls by several statisticians to update the introductory statistics curriculum to provide students with the computational tools and data-related capacity imperative for modern practice (e.g., Horton et al., 2014; Nolan & Temple Lang, 2010). Some instructors have already begun integrating more data and computational practices into the introductory course, including relational databases (e.g., Broatch et al., 2019), web scraping (e.g., Dogucu & Cetinkaya-Rundel, 2021), data wrangling (e.g., Hardin, 2018; McNamara & Horton, 2018), data cleaning (Holcomb & Spalsbury; 2005), multivariate visualization (e.g., Çetinkaya-Rundel & Tackett, 2020; Kaplan, 2018), reproducibility (e.g., Baumer et al., 2014), and version control tools (Beckman et al., 2021; Fiksel et al., 2019).

While these curricular innovations and the increased presence of data science, at least in department names (e.g., Rea, 2017), are positive, whether including computational practices in introductory statistics courses is becoming the norm or something that is only adopted by a handful of individual instructors is an open question. This paper introduces ***COM**putational **P**ractices in **U**ndergraduate **TE**aching of **S**tatistics* (COMPUTES), an instrument designed to measure the degree to which computational practices are being included in introductory statistics courses. This manuscript describes the development of COMPUTES and presents an analysis of its psychometric properties. Finally, instructors' responses are used to provide some empirical evidence about the degree to which computational practices are being introduced in the classroom.

## 2. INSTRUMENT DEVELOPMENT

To develop a blueprint for COMPUTES, the literature related to the teaching and learning of computation was reviewed. Much of this literature is focused at the K–12 level and predominantly addresses programming instruction (e.g., Carver & Klahr, 1986; Clements, 1991; Futschek & Moschitz, 2011; Marcelino et al., 2018; Sáez-López et al., 2016), or the development of computational reasoning (e.g., Barr & Stephenson, 2011; Benton et al., 2017; Brennan & Resnick, 2012). While this latter set of work helps provide a broad framework for big ideas and concepts that could be included in the curriculum, the components are often quite broad and hard to measure (e.g., abstraction, problem decomposition) or are focused on ideas that are unlikely to be included in introductory statistics courses (e.g., parallelization).

Weintrop et al. (2016) laid out a taxonomy underlying computational thinking in mathematics and science practices. These practices were identified through a comprehensive review of literature (related to computational thinking), classroom activities, and lesson plans and then validated through both interviews with scientists and mathematicians, and used with in-service high school teachers. The adopted framework categorizes computational thinking into four “distinct,...[yet] highly related and dependent” (Weintrop et al., 2016, p. 134) components: data practices, modelling and simulation practices, computational problem-solving practice, and systems thinking practices. Each component includes explicitly stated learning outcomes and details for mastery. The learning outcomes described aligned with much of the recommended content and many practices in introductory statistics courses.

This research used the taxonomy presented in Weintrop et al. (2016) as a framework to generate a blueprint for COMPUTES. Because the practices in the systems thinking category (e.g., defining systems and managing complexity) did not seem compatible with content typically included in introductory statistics courses, it was not included in the blueprint. The blueprint adopted included domains related to Data Practices, Modeling/Simulation Practices, and Computational Problem-Solving Practices. These three domains were minimally different from those presented in Weintrop et al. (2016).

## 2.1. ITEM GENERATION AND VALIDITY EVIDENCE

Fifty-one initial items were written based on the blueprint and, following the guidelines of Dillman et al. (2008), grouped together according to content: 20 Data Practices items, 14 Modeling/Simulation Practices items, 17 Computational Problem-Solving Practices items. Think-aloud interviews were conducted individually with three participants chosen because of their expertise in data science and statistics education. Each participant was administered the instrument and asked to articulate their reasoning as they responded to each item. When the participants had difficulty answering or understanding an item, they were probed for further thoughts. After completing the survey, the participants were asked if they thought the survey fit the current state of statistics education, was missing any important statistics education content, or contained unnecessary items.

These interviews resulted in improving item clarity and updating the instrument. First, additional items related to curricular emphasis on multivariate data were included in the Data Practices domain. Second, the items corresponding to the Simulation/Modeling domain seemed to be interpreted using the context of simulation-based inference, so this domain was relabeled Simulation Practices. Finally, most of the items written to address Computational Problem-Solving Practices were removed from the instrument. These items tended to be overly general and the think-aloud interviews made it clear these items were not being interpreted as intended. After reflection and discussion, the research team decided the majority of this domain was not well-suited for introductory statistics courses. The only items in the Computational Problem-Solving Practices that did seem pertinent were related to coding and debugging, so this domain was relabeled Coding Practices and all items not related to coding were omitted. The final instrument included 23 items measuring instructors’ curricular emphasis of computational practices across three domains: Data Practices (10 items), Simulation Practices (9 items), and Coding Practices (4 items). These items are presented in Appendix A.

## 2.2. RESEARCH QUESTIONS

Some degree of validity evidence is necessary in order to use results from COMPUTES to interpret results and draw inferences about computing practices in introductory statistics courses (American Educational Research Association et al., 2014). The cognitive interviews, described previously, are part of this trail of evidence, but are not sufficient to support a set of inferences. It is also important to undertake a psychometric analysis to understand the latent structure of the construct, item functioning, and score reliability (e.g., Borsboom et al., 2004). To that end, the following research questions are examined:

1. What is the latent structure underlying introductory statistics instructor’s responses to the COMPUTES items?
2. How well do the items fit the adopted model and how well do they measure the underlying construct(s)?

Finally, instructors' responses are used to provide empirical evidence about whether computational practices are being emphasized differently across institutional settings since these environments may have different resources/opportunities for their students (e.g., Kahlenberg, 2015). In particular:

3. Does the degree of curricular emphasis of computational practices vary by institutional setting (two-year colleges, four-year colleges, and universities)?

## 3. SAMPLE AND METHODS

The 23 items on COMPUTES were included as additional sections on the *Statistics Teaching Inventory* (STI; Zieffler et al., 2012) and were administered September and October of 2019 via Qualtrics. Email invitations were sent to five statistics education-focused listservs and mailing lists in the U.S.: ASA community of Isolated Statisticians (IsoStat), Consortium for the Advancement of Undergraduate Statistics Education (CAUSE), American Satistical Association Section on Statistics and Data Science Education, American Mathematical Association Two-Year Colleges, and the Mathematical Association of America Section on Statistics Education. These invitations solicited participation in the STI study for any instructors of a non-calculus-based introductory statistics course (e.g., courses aimed at non-quantitative majors, inference-based course) A total of 293 participants completed the STI, which resulted in 236 usable responses. Counts and proportions for the responses for each of the 23 items identified in the scale are presented in Table 1. Only one item, *understand concepts via simulation*, had a large number of non-response. Additionally, most items showed a pattern indicating the majority of respondents had little or no curricular emphasis on these ideas.

*Table 1. Counts (Proportions) of responses for each item*

| Item | None | Minor | Moderate | Major | NA |
|---|---|---|---|---|---|
| *Data Practices* | | | | | |
| Work with a codebook | 166 (0.70) | 44 (0.19) | 11 (0.05) | 7 (0.03) | 8 (0.03) |
| Use data stored in flat file | 58 (0.25) | 72 (0.31) | 52 (0.22) | 53 (0.22) | 1 (0.00) |
| Use data stored in relational database | 214 (0.91) | 12 (0.05) | 6 (0.03) | 3 (0.01) | 1 (0.00) |
| Collect data via web scraping | 208 (0.88) | 20 (0.08) | 5 (0.02) | 0 (0.00) | 3 (0.01) |
| Validate data | 133 (0.56) | 77 (0.33) | 18 (0.08) | 7 (0.03) | 1 (0.00) |
| Clean data | 127 (0.54) | 88 (0.37) | 15 (0.06) | 5 (0.02) | 1 (0.00) |
| Structure data | 157 (0.67) | 56 (0.24) | 15 (0.06) | 7 (0.03) | 1 (0.00) |
| Join datasets | 189 (0.80) | 29 (0.12) | 8 (0.03) | 2 (0.01) | 8 (0.03) |
| Produce visualizations of multivariate data | 111 (0.47) | 64 (0.26) | 34 (0.14) | 28 (0.11) | 6 (0.03) |
| Numerically summarize multivariate data | 139 (0.58) | 46 (0.19) | 21 (0.09) | 26 (0.10) | 7 (0.03) |
| *Simulation Practices* | | | | | |
| Identify real-world elements to include | 153 (0.65) | 38 (0.16) | 29 (0.12) | 13 (0.06) | 3 (0.01) |
| Decide what data will be produced | 149 (0.63) | 40 (0.17) | 30 (0.13) | 14 (0.06) | 3 (0.01) |
| Identify similarities/differences from real-world | 140 (0.59) | 40 (0.17) | 36 (0.15) | 17 (0.07) | 3 (0.01) |
| Describe impact of design on conclusions | 148 (0.63) | 40 (0.17) | 29 (0.12) | 17 (0.07) | 2 (0.01) |
| Understand concepts via simulation | 3 (0.01) | 23 (0.10) | 47 (0.20) | 47 (0.20) | 116 (0.49) |
| Evaluate conjecture about real-world | 138 (0.58) | 24 (0.10) | 31 (0.13) | 41 (0.17) | 2 (0.01) |
| Evaluate competing conjectures | 151 (0.64) | 30 (0.13) | 23 (0.10) | 30 (0.13) | 2 (0.01) |
| Generate data from model | 119 (0.50) | 83 (0.35) | 28 (0.12) | 6 (0.03) | 0 (0.00) |
| Generate data from sample | 133 (0.56) | 51 (0.22) | 25 (0.11) | 27 (0.11) | 0 (0.00) |
| *Coding Practices* | | | | | |
| Read code | 167 (0.71) | 23 (0.10) | 27 (0.11) | 18 (0.08) | 1 (0.00) |
| Modify code | 169 (0.72) | 14 (0.06) | 27 (0.11) | 25 (0.11) | 1 (0.00) |
| Debug code | 178 (0.75) | 39 (0.17) | 8 (0.03) | 10 (0.04) | 1 (0.00) |
| Create code | 186 (0.79) | 22 (0.09) | 14 (0.06) | 13 (0.06) | 1 (0.00) |

Two items—*relational databases* and *web scraping*, both in the Data Practices domain—were removed because the small amount of variation in instructor responses to these items prohibited estimation of the model parameters. All results are based on models fitted omitting these two items.

### 3.1. ANALYTIC METHOD

Instructors' response patterns were analyzed using multidimensional item response theory (MIRT). To account for the ordinal nature of the items, these models were fitted using a graded response model (GRM; Samejima, 1969). The GRM models the probability that person *i* endorses item *j* at level *k* conditional on the person's level of the underlying latent factor (θ) using a logistic model. Mathematically,

$$P(x_{ij} \geq k_j \mid \theta_i, \alpha_j, d_j) = \frac{1}{1 + exp\,[-D(\alpha_j^T \theta_i + d_j)]} \tag{1}$$

where $D$ is a scale factor (typically set to 1.7), $\alpha_j$ is an item discrimination parameter, and $d_j$ is a difficulty parameter for the *k*th response category of item *j*. All models were fitted to the data using the *mirt* package (Chalmers, 2012) in R.

To answer the first research question, three potential measurement models that might represent the underlying latent structure in the response data were evaluated. These three models are presented in Figure 1.

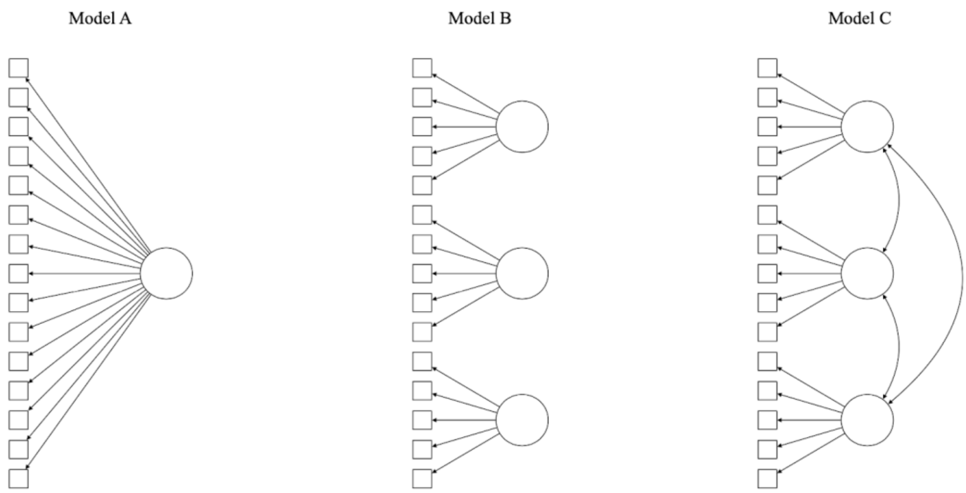


*Figure 1. Latent structure for three competing measurement models fitted to explain instructors' emphasis on computational reasoning.*

Model A posits that there is a single factor that explains all the covariance among items. In this model, the domains of Data Practices, Simulation Practices, and Coding Practices would not be unique, and instead form a single domain, say, general computational practices. In Model B, the covariance would be explained through three unique, yet uncorrelated, domains. The last candidate model, Model C assumes that the three domains are distinct, albeit correlated among one another.

To answer the second research question, the infit and outfit statistics for each of the items based on Model C were examined. These statistics help identify items and response patterns that have model–data misfit. The primary difference between these measures is that the infit measure is less sensitive to outlying item–person response patterns. (See de Ayala, 2009, for more technical detail.)

Items with perfect fit would have a mean square (MNSQ) value of 1. de Ayala suggests that items with MNSQ values "from 0.5 and 1.5 are 'okay' and those with values greater than 2 warrant closer inspection" (2009, p. 53). Both infit and outfit statistics can be standardized (ZSTD) and a *t*-test can be used to test the null hypothesis that there is perfect data–model fit. As such, items with “good” fit have ZSTD values between ±2 (e.g., Bond & Fox, 2001). With larger sample sizes, the recommendation is to examine the MNSQ values and if those are acceptable, ignore the ZSTD values (e.g., de Ayala, 2009).

To answer the last research question, estimates of the latent factors (i.e., factor scores) were computed for the sample of instructors and the distributions of these scores were compared across institutional types. Of the 236 respondents, 54 were instructors at two-year colleges, 87 were instructors at four-year colleges, and 87 were instructors at universities. Eight respondents classified their institution as “Other”. The factor scores for these respondents were not included in this part of the analysis.

The estimates of the latent factors were obtained using the observed item responses for each respondent, and the respective item parameters to compute expected a-posteriori (EAP) sum-scores. This assumed a multivariate Gaussian distribution for the priors (common assumption in applied IRT work), and quasi-Monte Carlo integration with 5000 nodes was used to determine the EAP estimates. (For more technical detail, see, e.g., Chalmers, 2016; Thissen et al., 1995)

## 4. RESULTS

In this section, the results of the analyses are presented for each research question.

### 4.1. RESEARCH QUESTION 1: WHAT IS THE LATENT STRUCTURE UNDERLYING INTRODUCTORY STATISTICS INSTRUCTOR’S RESPONSES TO THE COMPUTES ITEMS?

Information criteria (AICc, BIC, and SABIC) were computed for each of the competing models (Table 2). Based on these indices, Model C had the most empirical evidence given the data and the candidate models. This suggests the latent structure underlying instructors’ responses to the COMPUTES items is best captured using a correlated multidimensional model.

*Table 2. Information criteria for the three competing measurement models*

| Fit index | Model A | Model B | Model C |
|---|---|---|---|
| AICc | 8180.5 | 7384.7 | 7317.0 |
| BIC | 8376.9 | 7581.1 | 7514.9 |
| SABIC | 8110.6 | 7314.8 | 7239.1 |

Figure 2 shows the correlation structure between the three domains from Model C. This suggests the Data Practices and Coding Practices are quite related, and the Simulation Practices factor might be more distinct. To evaluate this, a correlated two factor model that combined the items from Data and Coding Practices into a single factor was also fitted. The fit indices (AICc = 7417.1; BIC = 7614.0; SABIC = 7344.6) suggest that this model does not fit better than the three-factor correlated multidimensional model (Model C). Thus, for the remainder of the article, we will use Model C to answer the other two research questions.

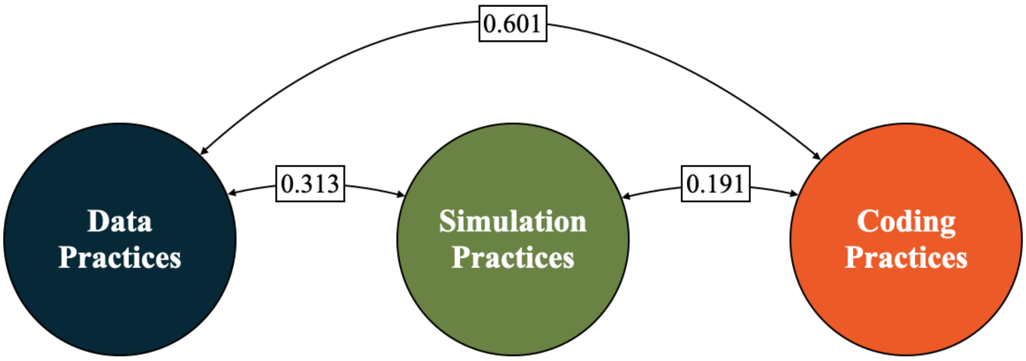


*Figure 2. Correlation structure among the three latent factors*

## 4.2. RESEARCH QUESTION 2: HOW WELL DO THE ITEMS FIT THE ADOPTED MODEL AND HOW WELL DO THEY MEASURE THE UNDERLYING CONSTRUCT(S)

The infit and outfit values for each item are presented in Table 3.

*Table 3. Item fit statistics for Model C. Statistics indicating item misfit have been bolded*

| | Outfit | | Infit | |
|---|---|---|---|---|
| Item | MNSQ | ZSTD | MNSQ | ZSTD |
| *Data Practices* | | | | |
| Work with a codebook | 0.72 | -0.70 | 0.93 | -0.28 |
| Use data stored in flat file | 0.88 | -1.08 | 0.87 | -1.20 |
| Validate data | 0.79 | -0.81 | 0.92 | -0.41 |
| Clean data | 0.69 | -1.39 | 0.78 | -1.25 |
| Structure data | 0.53 | -0.52 | 0.72 | -1.51 |
| Join datasets | 0.54 | -0.73 | 0.83 | -0.64 |
| Produce visualizations of multivariate data | 0.75 | -1.55 | 0.81 | -1.45 |
| Numerically summarize multivariate data | 0.70 | -1.34 | 0.80 | -1.39 |
| *Simulation Practices* | | | | |
| Identify real-world elements to include | 0.76 | -0.45 | 0.89 | -0.75 |
| Decide what data will be produced | 0.79 | -0.46 | 0.93 | -0.44 |
| Identify similarities/differences from real-world | 1.02 | 0.17 | 0.94 | -0.40 |
| Describe Impact of design on conclusions | 0.81 | -0.35 | 0.89 | -0.70 |
| Understand concepts via simulation | 0.87 | -0.91 | 0.90 | -0.72 |
| Evaluate conjecture about real-world | 0.64 | -0.90 | 0.76 | -1.70 |
| Evaluate competing conjectures | 0.77 | -0.29 | 0.85 | -1.01 |
| Generate data from model | 0.98 | -0.11 | 1.01 | 0.13 |
| Generate data from sample | 0.88 | -0.74 | 0.92 | -0.69 |
| *Coding Practices* | | | | |
| Read code | *0.04* | *-4.03* | *0.32* | *-2.35* |
| Modify code | 0.94 | 0.43 | 1.13 | 0.58 |
| Debug code | *0.21* | -1.96 | 0.72 | -1.03 |
| Create code | 0.41 | -0.07 | 1.05 | 0.27 |

In examining the MNSQ values, two items show MNSQ values under 0.5. Of those, only a single item in the Coding Practices domain (*read code*) has a ZSTD value less than –2. Turning to the infit statistics, one item (*read code*) had a MNSQ value under 0.5 and a ZSTD value less than –2. Since de Ayala (2009) suggested items with MNSQ values under 0.5 are not degrading to the measurement system, albeit less productive, this item was retained in the remainder of the analyses.

***Item parameters.*** Next the estimated item parameters from Model C were examined and are presented in Table 4. The *a*- and *d*-parameters give the estimated discrimination and threshold values for the fitted model. Table 4 displays the *a*- parameter for each domain in a single column.

*Table 4. Item parameters for Model C*

| | | | | | | MDIFF | | |
|---|---|---|---|---|---|---|---|---|
| *Item* | a | d1 | d2 | d3 | MDISC | 1 | 2 | 3 |
| *Data Practices* | *a1* | | | | | | | |
| Work with a codebook | 1.55 | -1.40 | -3.34 | -4.52 | 1.55 | 0.91 | 2.16 | 2.92 |
| Use data stored in flat file | 1.01 | 1.33 | -0.31 | -1.53 | 1.01 | -1.31 | 0.30 | 1.51 |
| Validate data | 1.92 | -0.52 | -3.34 | -5.08 | 1.92 | 0.27 | 1.74 | 2.64 |
| Clean data | 2.10 | -0.34 | -3.86 | -5.77 | 2.10 | 0.16 | 1.84 | 2.75 |
| Structure data | 3.54 | -1.94 | -5.38 | -7.72 | 3.54 | 0.55 | 1.52 | 2.18 |
| Join datasets | 1.76 | -2.34 | -4.29 | -6.19 | 1.76 | 1.33 | 2.44 | 3.52 |
| Produce visualizations of multivariate data | 1.45 | 0.13 | -1.50 | -2.81 | 1.45 | -0.09 | 1.04 | 1.94 |
| Numerically summarize multivariate data | 1.44 | -0.62 | -1.99 | -2.92 | 1.44 | 0.43 | 1.38 | 2.03 |
| *Simulation Practices* | *a2* | | | | | | | |
| Identify real-world elements to include | 3.74 | -3.10 | -5.79 | -8.01 | 3.74 | 0.83 | 1.55 | 2.14 |
| Decide what data will be produced | 4.16 | -3.16 | -6.32 | -8.64 | 4.16 | 0.76 | 1.52 | 2.08 |
| Identify similarities/differences from real-world | 3.43 | -1.98 | -4.42 | -6.96 | 3.43 | 0.58 | 1.29 | 2.03 |
| Describe impact of design on conclusions | 3.93 | -2.80 | -5.60 | -7.82 | 3.93 | 0.71 | 1.42 | 1.99 |
| Understand concepts via simulation | 1.82 | 3.21 | -0.29 | -2.67 | 1.82 | -1.76 | 0.16 | 1.46 |
| Evaluate conjecture about real-world | 3.84 | -2.11 | -3.83 | -5.92 | 3.84 | 0.55 | 1.00 | 1.54 |
| Evaluate competing conjectures | 3.88 | -2.98 | -5.01 | -6.63 | 3.88 | 0.77 | 1.29 | 1.71 |
| Generate data from model | 0.80 | -0.18 | -2.21 | -4.17 | 0.80 | 0.23 | 2.77 | 5.23 |
| Generate data from sample | 1.18 | -0.78 | -2.52 | -3.71 | 1.81 | 0.43 | 1.39 | 2.05 |
| *Coding Practices* | *a3* | | | | | | | |
| Read code | 16.43 | -6.76 | -14.34 | -27.69 | 16.43 | 0.41 | 0.87 | 1.68 |
| Modify code | 4.87 | -2.76 | -4.10 | -6.81 | 4.87 | 0.57 | 0.84 | 1.40 |
| Debug code | 5.82 | -4.15 | -9.45 | -10.81 | 5.82 | 0.71 | 1.62 | 1.86 |
| Create code | 3.85 | -3.47 | -5.38 | -6.95 | 3.85 | 0.90 | 1.39 | 1.80 |

The MDISC parameters, akin to the item discrimination parameters in a unidimensional IRT model, indicate how well the item discriminates among instructors of different levels of the latent factor—higher values are indicative of stronger relationships with the latent factor. (In a factor analytic perspective, these are akin to factor loadings, but they are not scaled between –1 and +1.) The estimated MDISC values were positive for all items suggesting the amount of agreement on each item is positively related to each of the latent factors. All items except one (*structure data*) in the Data Practices factor/domain had MDISC values between 0.5 and 3, values typically associated with "good" measurement (e.g., Baker, 2001; de Ayala, 2009). The items on the other two domains had MDISC values higher than 3 which may indicate these items over-discriminate on the latent factors; this is especially true for items in Coding Practices.

The estimated MDIFF values indicate the level of the latent factor at which a respondent is likely to transition to the next higher level of emphasis. More specifically, it reflects the level of the latent factor at which a respondent has a 0.5 probability of responding at the *k*th level of the item. In general, the thresholds should span the range of the latent variable that you intend to measure, in this case between –3 and +3. The MDIFF values for most of the items were positive, suggesting that the items provide better measurement for instructors who are above average on the latent factor. In fact, only three items seem to be measuring well for instructors who are below average on the latent factor.

***Reliability.*** Finally, estimates of the "marginal" score reliability for each of the measured domains were computed based on empirical estimates of the factor scores (Chalmers, 2012). These estimates give a sense of the proportion of the variance that is true score variance. For the measures, these were: 0.795 (Data Practices), 0.836 (Simulation Practices), and 0.703 (Coding Practices). The score reliability for each of these measures was reasonably high. Factor scores from the Coding Practices domain had the lowest reliability, probably due to the smaller number of items in the scale.

### 4.3. RESEARCH QUESTION 3: DOES THE DEGREE OF CURRICULAR EMPHASIS OF COMPUTATIONAL PRACTICES VARY BY INSTITUTIONAL SETTING (TWO-YEAR COLLEGES, FOUR-YEAR COLLEGES, AND UNIVERSITIES)?

Because of the exploratory nature of this part of the work, only a graphical analysis of these results is presented. Figure 3 shows the distribution of factor scores for each domain conditioned on institution type. The shape of the distributions for all three practices look somewhat similar across institution type and indicate there was variation in the instructional emphasis of these practices within each type of institution. The distributions for four-year college instructors, however, seem slightly shifted to the right relative to the other two institutions. This might indicate that a higher percentage of instructors at four-year colleges emphasize these practices.

Also of note, the distributions for the Simulation Practices domain seem bimodal, suggesting that there are subsets of teachers in all institution types who emphasize simulation practices in the curriculum and those who do not. This bimodality is also seen in the Coding Practices domain, but is more pronounced. Finally, there is almost no variation in emphasis of coding practices for two-year college instructors.

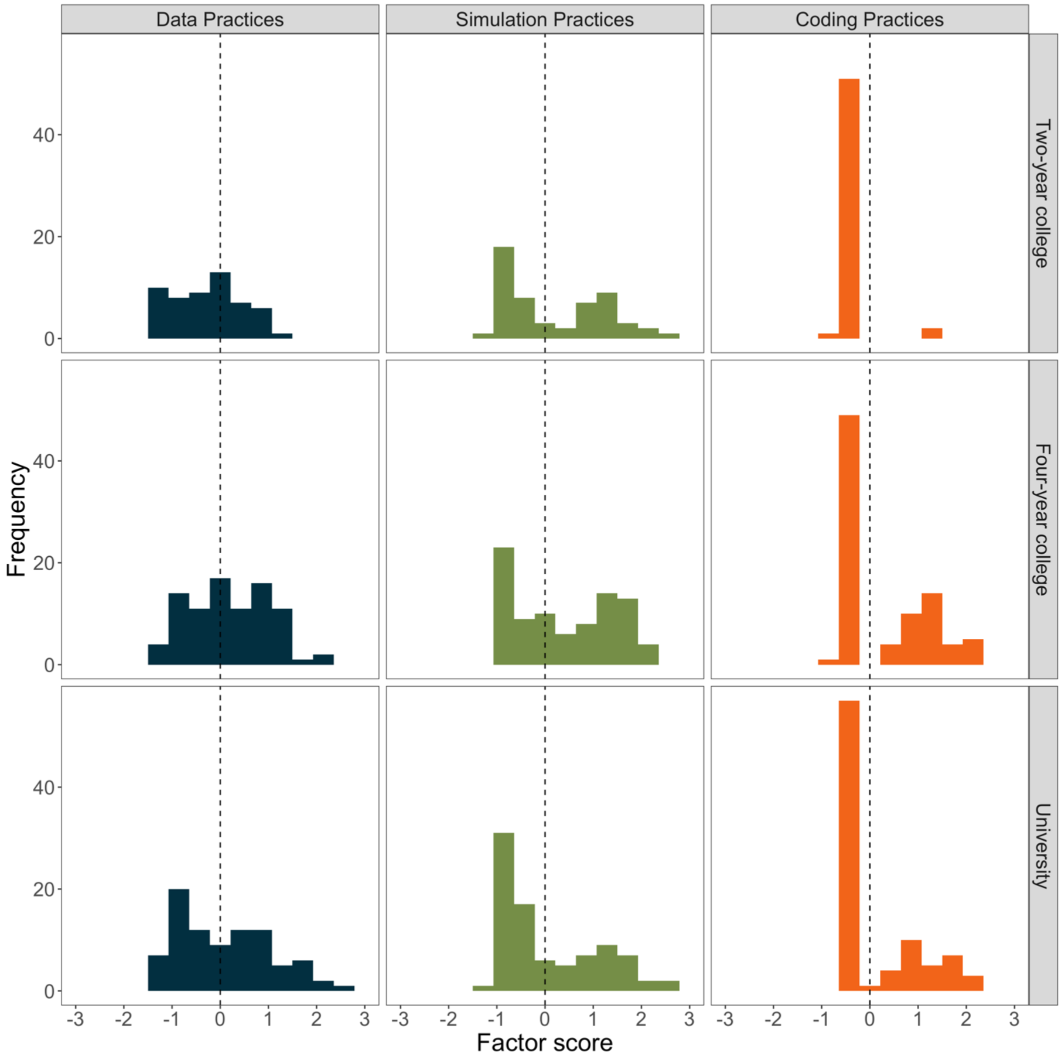


*Figure 3. Factor scores for the sample of instructors for each of the domains conditioned on institution type; Data Practices (left), Simulation Practices (middle), and Coding Practices (right). The vertical dashed line at 0 indicates the average amount of the latent factor*

To explore whether instructors emphasize computational practices more generally or focus on a single practice, the interrelationships among the factor scores were also examined. Because the dichotomy in coding practices was so pronounced, instructors were dichotomized based on their coding practices factor score using a cut-point of 0, and the scatterplot of the factor scores for the Simulation Practices domain versus the Data Practices domain for these two groups was examined (see Figure 4).

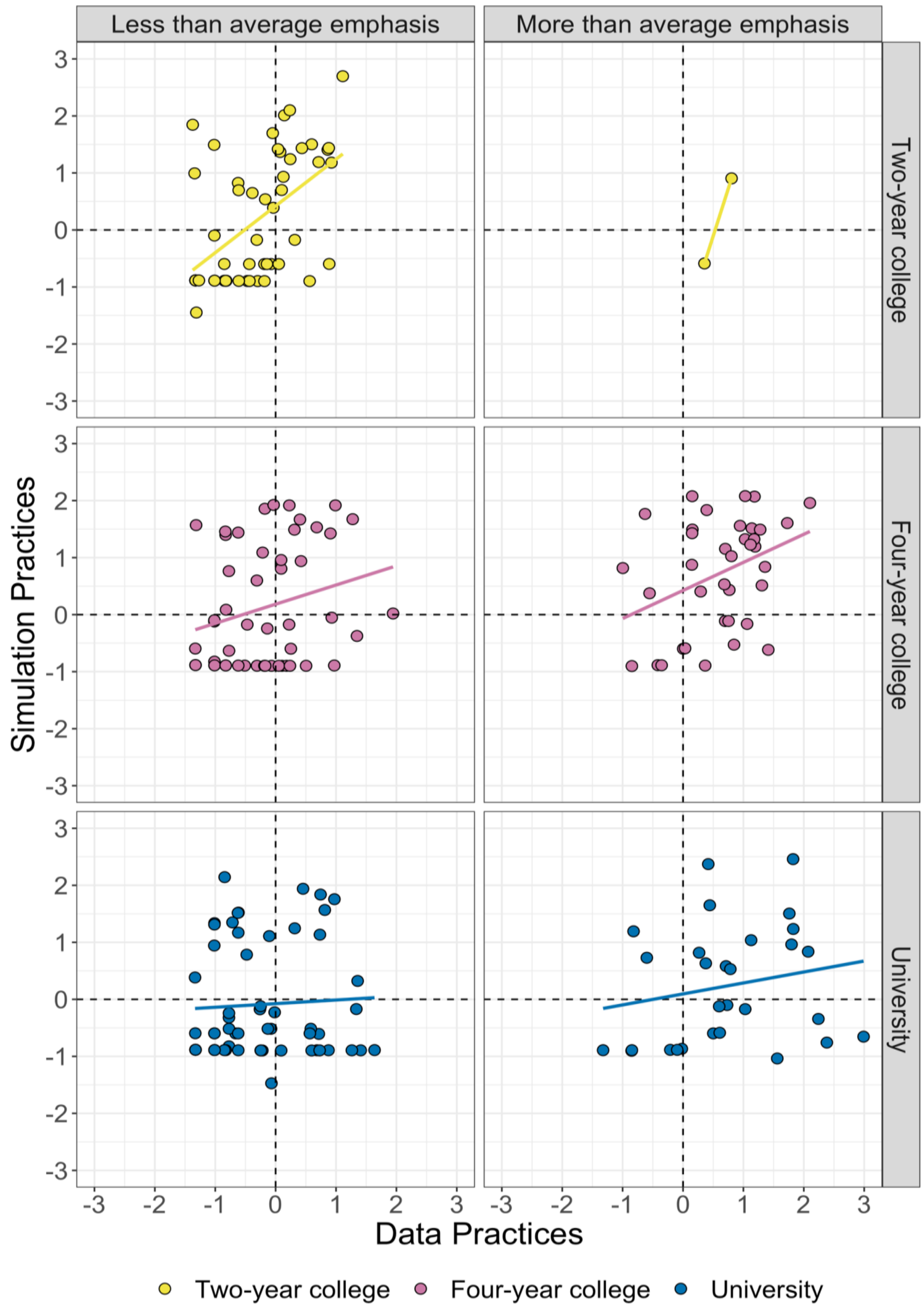


*Figure 4. Relationship between the factor scores for the Simulation Practices domain versus the Data Practices domain for the sample of instructors. This is conditioned on institution type and whether the instructor's factor score indicated they had below average emphasis on Coding Practices (left) or above average emphasis on Coding Practices (right). The fitted regression line is also shown to help discern the relationship. The horizontal and vertical dashed lines indicate an average amount of emphasis across institutions on the Simulation and Data Practices domains, respectively.*

From this plot it can be seen that instructors who had higher levels of emphasis on data practices also seem to have had higher levels of emphasis on simulation practices, regardless of whether or not they emphasized coding pracices. (The only exception to this seems to be university instructors who had a lower than average emphasis on coding practices.) The magnitude of this relationship seems to be higher for instructors who had above average emphasis on coding practices for both four-year college and university instructors. (With only two cases, we cannot conjecture whether this was also true for two-year college instructors.) These relationships are also consistent with estimated correlation structure (shown in Figure 2).

## 5. DISCUSSION

In this paper, the development of an instrument that can be used to provide empirical evidence about the degree to which computational practices are being emphasized in the classroom was described. Results from psychometric analyses were provided as evidence of validity to support the use of scores from the instrument. This research study specifically set out to answer three research questions: (1) What is the latent structure underlying introductory statistics instructor's responses to the COMPUTES items? (2) How well do the items fit the adopted model and how well do they measure the underlying construct(s)? And (3) Does the degree of curricular emphasis of computational practices vary by institutional setting (two-year colleges, four-year colleges, and universities)?

Based on the fit indices, responses to the COMPUTES items are multidimensional, albeit correlated, in nature. This is in line with Weintrop et al.'s (2016) statement that the domains are distinct, but related. The correlation structure among the latent factors suggests that instructors' emphasis on data practices and coding practices are highly related, while emphasis on simulation practices are less related to these two other domains. This may be because, in introductory statistics courses, simulation is seen (and subsequently taught) as a vehicle for statistical inference rather than employed as a computational practice. If this is the case, the Simulation Practices domain would be expected to have a higher correlation with the other two domains if the instrument is given to instructors teaching higher-level statistics courses.

The psychometric analysis suggested the items were generally productive (or at least not harmful) to the measurement process. The score reliability values for all three domains were reasonably high, suggesting the scores could be used in statistical analyses. The Coding Practices domain had the lowest score reliability, which is attributed to the small number of items.

The third research question examined whether computational practices are being emphasized differently across institutional settings (two-year colleges, four-year colleges and universities) since these environments may have different resources or opportunities for students. While there may be differences across institutional settings, these are generally slight. While not overcommitting to the small nuances between distributions given the small sample size (i.e., these results should be taken as hypothesis generation), there are some patterns that seem to stand out. The bimodality in the factor scores seen in the Coding Practices domain points toward a potential dichotomy across instructional settings. This may be due to a difference in curricular emphasis of the course content (e.g., a traditional inference-based course versus a data science-based course).

### 5.1. LIMITATIONS

There are several limitations to the research. The sampling method no doubt resulted in a biased sample. The listservs used to recruit participants probably drew more reform-oriented. While it seems reasonable to assume that reform-oriented instructors are more likely to emphasize computational practices related to simulation than their non-reform-oriented peers, it is unclear whether reform-oriented instructors have different curricular emphases on computational practices related to data or coding practices. A second related limitation is the sample size. While all the models converged, $N$=236 is somewhat small given the complexity of these models. Moreover, this sample was composed of instructors entirely based in the United States.

Another limitation is related to the content covered by the instrument. It may be that by omitting the two items (*relational databases* and *web scraping*) the remaining items do not adequately measure the instructors' curricular emphasis on data practices. It is noted that instructors' responses on one of the remaining items (*join datasets*) was highly correlated with responses on the relational databases item, so these items may be measuring the same thing.

Lastly, the differences in the factor score distributions are only exploratory. The same sample used to evaluate the instrument was also used in this analysis, which may bias the results. Hopefully, a more detailed analysis will be carried out on a larger sample of instructors in the future.

## 5.2. IMPLICATIONS AND FUTURE RESEARCH

Although the psychometric analysis suggested that the items generally measured the latent constructs, it also pointed toward ways to refine the COMPUTES instrument. The item thresholds were almost all above zero. This is problematic in that the instrument may have more difficulty discriminating among instructors who have little of the latent factor. This suggests that the instrument would benefit from additional items that measure differences at the lower end of the constructs.

The analyses also suggested the items in Coding Practices should be refined. The item statistics identified the *read syntax* item as a “bad” item, and pointed to the need for more nuanced items related to instructional emphasis on coding practices. For example, on reflection, perhaps the Coding Practices domain was too closely aligned with practices more prevalent in syntax-driven software. Future iterations of the instrument should include items that measure computational practices more likely to be included in courses that use software with a GUI (e.g., applying data moves; Erickson et al., 2019).

There is also much this instrument does not tell us and additional work that needs to be undertaken. For example, as pointed out by a reviewer, “although computational practices are recommended by many curriculum guidelines, an instructor may choose not to use computational practices based upon the needs of their particular students or because those computational practices are covered in other courses.” Understanding why instructors are not including computational practices for particular audiences or in certain courses would help teachers and curriculum designers think about whether and how to include additional computational practices. This could also be useful for considering how ideas/concepts from computation might be introduced in courses that do not emphasize computation (e.g., in a statistical literacy course). The AP Computer Science Principles course (College Board, 2020) might be a good model for this.

It would be valuable to administer COMPUTES to instructors outside the United States. Given some of the international efforts in addressing data and computation, it would be interesting to see if there are differences in the types of computational and data practices being emphasized across countries. This might point to ideas that could be adopted or modified to expand computing’s role in the curriculum more generally and for a more diverse audience.

Another opportunity for additional insight is to study the efficacy of the teaching of computational and data practices. While including computational and data practices in the curriculum is a necessary condition for effective teaching of these ideas, it is not sufficient. Case studies of classroom teaching practices might be one way to begin to examine this. Ultimately, however, it will be important to assess and understand the extent to which students have learned these computational and data practices.

It is exciting to have an instrument that can be used to measure curricular emphasis of computational practices. After adding and modifying some items, the COMPUTES instrument could be used to evaluate instructional changes in computational practices. As such, it could be used to evaluate how professional development (e.g., workshops) impact curricular change. Lastly, given the increased role of computation in the 21st century workforce, it is possible that the results from future administrations of COMPUTES could be used to call attention to potential disparities in students’ learning opportunities.

## REFERENCES

American Educational Research Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association American Psychological Association National Council on Measurement in Education.

American Statistical Association. (2017). *Curriculum guidelines for undergraduate programs in statistical science*. American Statistical Association.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed., ERIC Document Reproduction Service No. ED 458 219). Eric Clearing House on Assessment and Evaluation.

Barr, B. V., & Stephenson, C. (2011). Bringing computational thinking to K–12: What is involved and what is the role of the computer science education community? *ACM Inroads*, *2*(1), 48–54. https://doi.org/10.1145/1929887.1929905

Baumer, B. S., Cetinkaya-Rundel, M., Bray, A., Loi, L., & Horton, N. J. (2014). R Markdown: Integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education, 8*(1). https://escholarship.org/uc/item/90b2f5xh

Beckman, M. D., Çetinkaya-Rundel, M., Horton, N. J., Rundel, C. W., Sullivan, A. J, & Tackett, M. (2021). Implementing version control with Git and GitHub as a learning objective in statistics and data science courses. *Journal of Statistics and Data Science Education, 29*(sup1), S132-S144. https://doi.org/10.1080/10691898.2020.1848485

Benton, L., Hoyles, C., Kalas, I., & Noss, R. (2017). Bridging primary programming and mathematics: Some findings of design research in England. *Digital Experiences in Mathematics Education*, *3*(2), 115–138. https://doi.org/10.1007/s40751-017-0028-x

Blair, R., Kirkman, E., & Maxwell, J. (2018). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2015 CBMS Survey*. American Mathematical Society. https://doi.org/10.1090/cbmssurvey/2015

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model* (2nd ed.). Lawrence Erlbaum.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Brennan, K. & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In A. F. Ball & C. F. Tyson (Eds.), *Non satis scire: To know is not enough. Proceedings of the 2012 Annual Meeting of the American Educational Research Association*, April 13–17, Vancouver, Canada (Vol. 1, pp. 1–25). American Educational Research Association.

Broatch, J. E., Dietrich, S., & Goelman, D. (2019). Introducing data science techniques by connecting database concepts and dplyr. *Journal of Statistics Education*, *27*(3), 147–153. https://doi.org/10.1080/10691898.2019.1647768

Business Higher Education Forum and PricewaterhouseCoopers. (2017), *Investing in America's data science and analytics talent: The case for action.* http://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf

Carver, S. M., & Klahr, D. (1986). Assessing children's LOGO debugging skills with a formal model. *Journal of Educational Computing Research*, *2*(4), 487–525. https://doi.org/10.2190/KRD4-YNHH-X283-3P5V

Çetinkaya-Rundel, M., & Tackett, M. (2020). From drab to fab: Teaching visualization via incremental improvements. *CHANCE*, *33*(2), 31–41. https://doi.org/10.1080/09332480.2020.1754074

Chalmers, R. P. (2016). Generating adaptive and non-adaptive test interfaces for multidimensional item response theory applications. *Journal of Statistical Software*, *71*(5), 1–39. http://dx.doi.org/10.18637/jss.v071.i05

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Clements, D. H. (1991). Enhancement of creativity in computer environments. *American Educational Research Journal*, *28*(1), 173–187. https://doi.org/10.2307/1162883

Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, *69*(4), 266–282. https://doi.org/10.1080/00031305.2015.1093029

College Board. (2020). AP computer science principles: Course and exam description. Author. https://apcentral.collegeboard.org/courses/ap-computer-science-principles/course

de Ayala, R. J. (2009). *The theory and practice of item response theory*. The Guilford Press.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2008). *Internet, mail, and mixed-mode surveys: The tailored design method*. John Wiley.

Dogucu, M., & Çetinkaya-Rundel, M. (2021). Web scraping in the statistics and data science curriculum: Challenges and opportunities. *Journal of Statistics Education, 29*(supp1), S112–S122. https://doi.org/10.1080/10691898.2020.1787116

Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal*, *16*(1), 44–49. https://doi.org/10.52041/serj.v16i1.213

Engel, J., Biehler, R., Frischemeier, D [Daniel], Podworny, S., Schiller, A., & Martignon, L. (2019). *Zivilstatistik: Konzept einer neuen Perspektive auf Data Literacy und Statistical Literacy. AStA Wirtschafts- und Sozialstatistisches Archiv*, *13*(3–4), 213–244. https://doi.org/10.1007/s11943-019-00260-w

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data moves. *Technology Innovations in Statistics Education*, *12*(1). https://escholarship.org/uc/item/0mg8m7g6

Fiksel, J., Jager, L. R., Hardin, J. S., & Taub, M. A. (2019). Using gitHub classroom to teach statistics. *Journal of Statistics Education*, 27(2), 1–10. https://doi.org/10.1080/10691898.2019.1617089

Forbes, S. (2014). The coming of age of statistics education in New Zealand, and its influence internationally. *Journal of Statistics Education*, *22*(2), 2. https://doi.org/10.1080/10691898.2014.11889699

Forbes, S., Chapman, J., Harraway, J., Stirling, D., & Wild, C. (2014). Use of data visualization in the teaching of statistics: A New Zealand perspecitve. *Statistics Education Research Journal, 13*(2), 187–201. https://doi.org/10.52041/serj.v13i2.290

Futschek, G., & Moschitz, J. (2011). Learning algorithmic thinking with tangible objects eases transition to computer programming. In I. Kalaš & R.T. Mittermeir (Eds.), *Situation, evolution, and perspectives. Proceedings of the 5th International Conference on Informatics in Schools*, Bratislava, Slovakia, October 26–29, (pp. 155–164). Springer.

Gal, I., Ridgway, J., & Nicholson, J. (2016, December). Exploration of skills andconceptual knowledge needed for understanding statistics about society: A workshop. In J. Engel (Eds.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE)*, Berlin, Germany. https://iase-web.org/documents/papers/rt2016/Gal_Workshop.pdf?1482483777

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal*, *16*(1), 22–25. https://doi.org/10.52041/serj.v16i1.209

Hardin, J. (2018). Dynamic data in the statistics classroom. *Technology Innovations in Statistics Education, 11*(1). https://escholarship.org/uc/item/13g5g3dm

Hardy, Q. (2012, January 30). BITS; The popularity of statistics. *New York Times* [online]. http://query.nytimes.com/gst/fullpage.html?res=9B01EED61538F933A05752C0A9649D8B63

Holcomb, J. P., & Spalsbury, A. (2005). Teaching students to use summary statistics and graphics to clean and analyze data. *Journal of Statistics Education*, *13*(3). https://doi.org/10.1080/10691898.2005.11910567

Holdren, J. P., & Lander, E. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. President's Council of Advisors on Science and Technology. https://eric.ed.gov/?id=ED541511

Horton, N. J., Baumer, B. S., & Wickham, H. (2014). Teaching precursors to data science in introductory and second courses in statistics. *arXiv:*1401.3269 [stat.CO]. https://arxiv.org/

Horton, N. J., & Hardin, J. S. (2015). Teaching the next generation of statistics students to "think with data". *The American Statistician*, *69*(4), 259–265. https://doi.org/10.1080/00031305.2015.1094283

IDSSP Curriculum Team (2019). *Curriculum Frameworks for Introductory Data Science*. http://idssp.org/files/IDSSP_Frameworks_1.0.pdf

Kahlenberg, R. D. (2015). *How higher education funding shortchanges community colleges*. The Century Foundation. https://vtechworks.lib.vt.edu/handle/10919/83636

Kaplan, D. (2018). Teaching stats for data science. *The American Statistician, 72*(1), 89–96. https://doi.org/10.1080/00031305.2017.1398107

Marcelino, M. J., Pessoa, T., Vieira, C., Salvador, T., & Mendes, A. J. (2018). Learning computational thinking and scratch at distance. *Computers in Human Behavior*, *80*, 470–477. https://doi.org/10.1016/j.chb.2017.09.025

McNamara, A., & Horton, N. J. (2018). Wrangling categorical data in R. *The American Statistician*, *72*(1), 97–104. https://doi.org/10.1080/00031305.2017.1356375

Melguizo, T., & Wolniak, G. C. (2012). The earnings benefits of majoring in STEM fields among high achieving minority students. *Research in Higher Education, 53*(4) 383–405. https://doi-org.ezp1.lib.umn.edu/10.1007/s11162-011-9238-z

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options.* The National Academies Press. https://doi.org/10.17226/25104

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician, 64*(2), 97–107. https://doi.org/10.1198/tast.2010.09132

Rash, S. (Director). (1987). *Can't buy me love* [Film]. Touchstone Pictures.

Rea, S (2017, October 16). Carnegie Mellon changes statistics department's name to reflect its eminent position in data science research and education. https://www.cmu.edu/dietrich/news/news-stories/2017/october/stats-name-change.html#:~:text=Carnegie%20Mellon%20University%20has%20changed,Department%20of%20Statistics%20%26%20Data%20Science

Reckase, M. D. (2009). *Multidimensional item response theory*. Springer. https://doi.org/10.1007/978-0-387-89976-3

Ridgway, J. (2016). Implications of the data revolution for statistics education: The data revolution and statistics education. *International Statistical Review*, *84*(3), 528–549. https://doi.org/10.1111/insr.12110

Ridgway, J., Ridgway, R., & Nicholson, J. (2018, March). Data science for all: A stroll in the foothills. In M. A. Sorto (Ed.), *Looking back, looking forward*. *Proceedings of the 10th International Conference on the Teaching of Statistics (ICOTS10),* Kyoto, Japan, July 8–14. International Statistical Institute.

Sáez-López, J. M., Román-González, M., & Vázquez-Cano, E. (2016). Visual programming languages integrated across the curriculum in elementary school: A two year case study using "Scratch" in five schools. *Computers & Education*, *97*, 129–141. https://doi.org/10.1016/j.compedu.2016.03.003

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17, Part 2). Psychometric Society. http://dx.doi.org.ezp1.lib.umn.edu/10.1007/BF03372160

Schiller, A., & Engel, J. (2016). Civic statistics and the preparation of future secondary school mathematics teachers. In J. Engel (Eds.), *Promoting understanding of statistics about society: Proceedings of the Roundtable Conference of the International Association of Statistics Education*, Berlin, Germany. https://iase-web.org/documents/papers/rt2016/Schiller.pdf

Thissen, D., Pommerich, M., Billeaud, K., & Williams, V. S. L. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement, 19*, 39-49. https://doi-org.ezp1.lib.umn.edu/10.1177/014662169501900105

U.S. Department of Commerce. (2017). Women in STEM: 2017 Update. ESA Issue Brief #06-17. https://www.commerce.gov/news/fact-sheets/2017/11/women-stem-2017-update

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://www.jstor.org/stable/43867736

Zieffler, A., Park, J., Garfield, J., Delmas, R., & Bjornsdottir, A. (2012). The statistics teaching inventory: A survey on statistics teachers' classroom practices and beliefs. *Journal of Statistics Education*, *20*(1). https://doi.org/10.1080/10691898.2012.11889632

CHELSEY LEGACY
Education Sciences Building
56 E River Rd
Minneapolis, MN 55455

# APPENDIX A

All items were selected-response using a Likert scale of: No emphasis, minor emphasis, moderate emphasis, and major emphasis. Items with greyed out text were ultimately dropped from the scale.

**Data Practices**

1. How much emphasis is placed on having students work with a data codebook?
2. How much emphasis is placed on having students use data stored in a flat file (e.g., CSV, TXT, SAV)?
3. How much emphasis is placed on having students use data stored in a relational database (e.g., mySQL)?
4. How much emphasis is placed on having students collect data via web scraping?
5. How much emphasis is placed on having students validate data (e.g., range checking, variable type)?
6. How much emphasis is placed on having students clean data (e.g., error coding, recoding, duplicate case elimination)?
7. How much emphasis is placed on having students structure data (e.g., reshaping, filtering, subsetting)?
8. How much emphasis is placed on having students join/merge multiple datasets together?
9. How much emphasis is placed on having students produce visualizations of multivariate data with technology?
10. How much emphasis is placed on having students produce numerical summaries of multivariate data with technology?

**Simulation Practices Domain**

1. How much emphasis is placed on having students identify elements of the real-world phenomena that will be included in the simulation?
2. How much emphasis is placed on having students decide what data will be produced by the simulation?
3. How much emphasis is placed on having students identify similarities/differences between the simulation and the real-world phenomenon being simulated?
4. How much emphasis is placed on having students describe how the design of the simulation (e.g., assumptions, choices) impact the conclusions drawn?
5. How much emphasis is placed on having students use simulation to advance their understanding of statistical concepts through interacting with a simulation?
6. How much emphasis is placed on having students use simulation to evaluate a conjecture/claim about a real-world phenomenon?
7. How much emphasis is placed on having students use simulation to evaluate competing conjectures/claims about a real-world phenomenon?
8. How much emphasis is placed on having students generate data from a model (e.g., random sample from a Normal distribution)?
9. How much emphasis is placed on having students generate data from a sample (e.g., bootstrapping, randomizing)?

**Coding Practices**

1. How much emphasis is placed on having students read/understand code/syntax?
2. How much emphasis is placed on having students modify existing code/syntax?
3. How much emphasis is placed on having students debug code/syntax?
4. How much emphasis is placed on having students create code/syntax from scratch?