

Article

How the Mastery Rubric for Statistical Literacy Can Generate Actionable Evidence about Statistical and Quantitative Learning Outcomes

Rochelle E. Tractenberg

Collaborative for Research on Outcomes and Metrics; Departments of Neurology; Biostatistics, Bioinformatics & Biomathematics and Rehabilitation Medicine, Georgetown University Medical Center, Suite 207 Building D, 4000 Reservoir Road NW, Washington, DC 20057, USA; rochelle.tractenberg@gmail.com

Academic Editor: James Albright

Received: 7 July 2016; Accepted: 12 December 2016; Published: 24 December 2016

Abstract: Statistical literacy is essential to an informed citizenry; and two emerging trends highlight a growing need for training that achieves this literacy. The first trend is towards “big” data: while automated analyses can exploit massive amounts of data, the interpretation—and possibly more importantly, the replication—of results are challenging without adequate statistical literacy. The second trend is that science and scientific publishing are struggling with insufficient/inappropriate statistical reasoning in writing, reviewing, and editing. This paper describes a model for statistical literacy (SL) and its development that can support modern scientific practice. An established curriculum development and evaluation tool—the Mastery Rubric—is integrated with a new, developmental, model of statistical literacy that reflects the complexity of reasoning and habits of mind that scientists need to cultivate in order to recognize, choose, and interpret statistical methods. This developmental model provides actionable evidence, and explicit opportunities for consequential assessment that serves students, instructors, developers/reviewers/accreditors of a curriculum, and institutions. By supporting the enrichment, rather than increasing the amount, of statistical training in the basic and life sciences, this approach supports curriculum development, evaluation, and delivery to promote statistical literacy for students and a collective quantitative proficiency more broadly.

Keywords: statistical literacy; mastery rubric; collective quantitative proficiency; basic sciences; life sciences; scientific practice; curriculum development; curriculum evaluation; actionable evidence

1. Introduction

Statistical literacy (SL) is widely described as important for full social participation (see [1]; elementary curricula, e.g., [2,3]; higher education and beyond, e.g., [4–6]). Although this is true for all students, there is a special relationship between statistics and scientific research that amplifies the importance of developing appropriate statistical literacy in undergraduate or graduate/post-graduate students in the sciences.

Empirical research relies on statistical methods, and statistics is a wide, dynamic field perpetually propelled by new and improved methods. This far outstrips the capacities of other fields to fully adapt to these innovations, much less to incorporate all “relevant” methods in their own PhD curricula. Recently, Weissgerber et al. (2016) [7] correctly articulate that—and the myriad empirical arguments why—basic scientists *need* training in statistics (see also [8–16]; see also [17]). In fact, science PhD programs face a nearly Sisyphusian task: to adapt to some or any new methods, or even to prepare their students to adapt, so that their non-statistical discipline may exploit the power of new, or justify selecting established, statistical methods. Learning all statistical methods is clearly not feasible; even

a focus on “just” those that are currently relevant for the discipline may impede adoption of newer, more efficient methods in the future. However, initiating the development of statistical literacy and *orienting* science students to value quantitative methods, which empowers them to seek additional training when needed, might be an achievable goal.

Exemplifying the special importance of statistical literacy for scientists as they are trained is the Carnegie model of the doctorate wherein PhD programs prepare graduates to be/become “stewards” of their scientific disciplines (see [18] (pp. 9–14)). The definition of a steward of a discipline is “someone who will creatively generate new knowledge, critically conserve valuable and useful ideas, and responsibly transform those understandings through writing, teaching, and application” [18] (p. 5). Consistent with the disciplinary stewardship model, Henson et al. (2010) [12] propose a “collective quantitative proficiency” (CQP) model explicitly linking the valuation of quantitative methods within the culture of a scientific discipline to the training in these methods that is provided to the future researchers in (stewards of) that discipline. The CQP was described originally for education researchers, but the argument and model are appropriate to all sciences. In fact, Weissgerber et al. (2016) [7] review only the most recent literature representing the damage that weak or incomplete (or incorrect) knowledge of statistics and statistical methods is currently having on the rigor, interpretability and reproducibility of scientific work across the basic and life sciences. Established scientific practitioners must become more statistically literate to effectively model this competency for their mentees and students, to teach effectively, and to promote competence in writing, reviewing, and editing across the sciences. As Shulman noted, “(b)oth scholarship and teaching in any field reflect the character of inquiry, the nature of community, and the ways in which research and teaching are conducted in that particular discipline or disciplinary intersection” [19] (p. xii). Students at all levels need to know (and observe) that their scientific mentors also value—and contribute to—the collective quantitative proficiency (CQP) that disciplinary stewardship requires.

Despite the importance of statistical literacy and competency for the practicing scientist and the steward, doctoral programs may struggle with recommendations to *add* statistical training (see [20–22]). Many science PhD programs include *no formal statistical training*, or just a single course (see [23]; see also [7]). Two emerging trends in the basic and life sciences are highlighting a growing need for the addition—and integration- of statistical training in these disciplines. The first trend is towards “big” data across basic and life sciences; where the potential to automate—and thereby remove from active consideration—statistical inferences across datasets could ultimately exclude formal training and reasoning in statistics and experimental design. While some PhD programs contemplate adding statistical training to their programs, there is also movement to integrate “big data” into training future or current stewards of the biomedical sciences—*without* attention to reproducibility, experimental design, inferential statistics, or statistical literacy (e.g., [24]). While automated analyses can exploit massive amounts of data, without statistical literacy, the interpretation—and possibly more importantly, the replication, of results is challenged. However, “statistical literacy” is not included as a key competency in most fields (e.g., bioinformatics [25]; biology [26]) and where it is discussed, it relates to undergraduate single-course educational requirements (compliance) or to something less concretely defined (e.g., [27–29]). These arguments focus on undergraduate and PhD level programs because at the Master’s level, the course load is usually rigidly fixed; however, those seeking or completing Master’s level preparation are also challenged when it comes to statistical literacy.

It seems impossible to achieve the goal of a “collective quantitative proficiency” [12] among disciplinary stewards given the resistance to (or lack of time for, or lack of opportunities/interest in) coursework beyond introductory statistical training (e.g., [22]). However, adding or retaining one course in “introductory statistics” is also unlikely to achieve sufficient statistical literacy for modern scientific practice—as either a producer or a consumer of argument that relies on quantitation and data. A one-course approach to statistical literacy for PhD programs *implies* that:

- (A) the single course is sufficient to teach the critical—and complex—set of skills that encompasses “the ways in which research . . . (is) conducted in that particular discipline [19]; and

- (B) the single course will support the level of consumption and production of statistical arguments representing competent stewardship of a discipline that uses these methods.

The one-course-done model of statistical training exemplifies the comment by Henson et al. (2010) [12] (p. 235) that “(f)aculty and students often perceive quantitative methods as a static field to be mastered”. Moreover, the current conceptualizations of statistical literacy are grounded in the satisfaction of an undergraduate requirement (e.g., [30,31], for example, the Guidelines for Assessment and Instruction in Statistics Education (GAISE [31]; see also e.g., [6]). The scientist, professional, and/or instructor *must* be considered to have statistical literacy needs that differ qualitatively and quantitatively from those of undergraduates whose use for, or application of, statistical reasoning and methods *is not yet known*. For professional scientists, statistical literacy must support the responsible stewardship of their disciplines, producing and consuming statistical arguments (see e.g., [12]; [32] (p. xiii); see also [23]). This is a complex set of skills required for literature review, documenting the background and contextual (apart from the statistical) significance of one’s work, and for writing and reviewing manuscripts. Instead of reinforcing the perception that quantitative methods are “static”, an explicitly *developmental* model of statistical literacy directs attention of PhD scientists (students and mentors alike) towards their own awareness of the importance of, and variety in, quantitative method options for their research and discipline. Because the model is developmental, it can be augmented to accommodate learners *earlier* (than the PhD; see [33]) in their training. The model, described in the next section, is intended to:

- A. promote metacognitive awareness of what statistical literacy encompasses for disciplinary stewards;
- B. exemplify the link between this statistical literacy and the “collective quantitative proficiency” of Henson et al. (2010) [12]; and
- C. represent statistical literacy training that could be integrated into—or at least initialized within—any PhD science program (and possibly earlier).

This conceptualization of statistical literacy as developmental could fulfill the objectives of increasing statistical sophistication for scientists, reviewers, and faculty/mentors who are training future scientists, reviewers and faculty/mentors. Moreover, although other models have separated statistical “literacy”, “reasoning”, and “thinking” [34], these are actually three stages in a developmental trajectory that describes a deepening of sophistication with respect to data and principles of statistics. A more explicit statement of this development is intended to promote the “cultural” shift *towards* CQP in PhD training in basic and life sciences like biology, physiology, biochemistry, and genetics—towards a more holistic, reflective, and adaptive view of statistical literacy (SL). A curriculum development and evaluation tool, the Mastery Rubric (described in the next section), can be used to create, evaluate, or revise curricula that can generate *actionable evidence* (see [35]) of performance by students, instructors, and institutions. In this manuscript, a new Mastery Rubric for Statistical Literacy (MR-SL) is presented, and its potential to generate actionable evidence of growth and development in understanding of fundamental statistical concepts, and reasoning with them, is explored.

The Mastery Rubric

A traditional rubric is assignment-specific and lists the skills the grader requires in the work product, along with performance levels from poorest to best [36] (Chapter 1). The Mastery Rubric is similar, but outlines the knowledge, skills and abilities (KSAs) to be developed *within the curriculum* (or over time), together with performance levels that characterize the learner moving from novice to expert [37,38].

Related to the Mastery Rubric is the concept of a “learning progression” (e.g., [39] (p. 1)) which describes shifts from naïve to “more expert understanding” and is based on how children learn the

concepts of interest (but see [40] for an example with law students). Whereas a learning progression represents a curricular segment (e.g., Schwarz et al. 2009 [41]), the Mastery Rubric [37,42,43] represents the entire (predominantly) post-baccalaureate curriculum. Like the Mastery Rubric for Statistical Literacy (MR-SL, Table 2), two others were designed to capture and encourage development *throughout the career* [42,43]. Additionally, unlike a learning progression, the Mastery Rubric is public: explication of curricular objectives, and what work products look like then these are met, facilitates the identification by faculty, mentors, or evaluators of strengths and weaknesses in the curriculum itself. This also formalizes the evidence that any individual may elicit (instructor/institution) or present (student) to support their claim of achieving a target performance level throughout the curriculum. This can also support faculty in *other* courses to create opportunities to generate this evidence, and instruction supporting the same objectives from diverse contexts and perspectives. Explicit and public description of the necessary evidence can, in turn promote learners to self-monitor, and spur the individual (student, instructor, or institution) to seek (or create) opportunities to generate such evidence [42].

The Mastery Rubric represents the perspective of Messick (1994) [44]: articulating what KSAs students should possess at the end of the curriculum; what behaviours by the students will reveal these KSAs; and what tasks will elicit these specific behaviours. Toohey (1999) [45] refers to this outcomes-based approach as “systems-” or “performance-based”, and every Mastery Rubric follows this approach. Thus, by design, any Mastery Rubric supports assessable curriculum development, evaluation, and delivery because learning objectives are articulated and public so that each can be explicitly aligned to individuals’ progress and development along the articulated continuum from novice to expert. Then, conversations about curricular objectives, and actionable evidence of whether or not they are being met, are possible for all stakeholders.

In the next sections, the MR-SL is presented and described, and its alignment with principles of learning outcomes documentation [46] is analyzed.

2. Materials and Methods

Every Mastery Rubric is constructed with two dimensions: performance levels that represent a developmental trajectory (columns) and knowledge, skills, and abilities that represent the targets of the teaching and/or learning (rows; [38]). The methods by which each dimension of the MR-SL was constructed are articulated below. A degrees of freedom analysis [47–49] was used to create a matrix to permit examination of alignment of features of the MR-SL with the Principles for Learning Outcomes articulated by the National Institution for Learning Outcomes Assessment (NILOA [46]).

2.1. The Mastery Rubric for Statistical Literacy (MR-SL): Establishing a Developmental Trajectory

As noted, one of the two essential elements in the creation of a Mastery Rubric (MR) is the articulation of a developmental trajectory. Much of the research in statistical literacy has focused on understanding how students or experts think about data (e.g., [50–52])—which means that the two ends of the “developmental trajectory” in this discussion to date are “completing the undergraduate course” and “being an expert”.

The Mastery Rubric for Statistical Literacy (MR-SL) was designed from the *opposite* perspective, namely, to articulate what is common across middle stages of engagement with data (consumption and production), with desired entry and exit criteria for each stage, along an explicit continuum from more naïve to more expert. This is achieved by explicit reference to Bloom’s Taxonomy of Educational Objectives [53]; see also [54]. Moreover, the MR-SL was constructed synthesizing a developmental view of Bloom’s taxonomy with a long-standing model of the development of general literacy [55], focusing on the knowledge, skills, and abilities specific to statistical literacy arising by consensus from the literature (e.g., [50–52,56–58]).

Table 1 presents the Bloom’s Taxonomic context of the MR-SL.

Table 1. Performance levels in the developmental trajectory of “statistical literacy”: Given a research question, proposal, manuscript, report, or grant, this reader will/is.

Pre-Literate	Beginning Literacy	Functionally Literate	Skilled (Fluent)	Independent (Journeyman)	Expert (Master)
Read or skip stats/methods sections—no critique or evaluation. Assume writer (and/or publisher) must know what they’re doing. Accept results without question. Unengaged with statistical reasoning, lacking quantitative habits of mind or an awareness of their role in science.	Read, generally understand, notice gross errors, e.g., if categorical method applied to continuous variable or vice versa. Developing meta-cognitive awareness that if a question arises in their mind, the method may not be correct or clearly articulated. Initial engagement with statistical reasoning, developing awareness of this skill and how to grow/use it.	Consolidating reading and understanding, beginning to learn how to analyze (with software). Awareness of rules of thumb (e.g., sample size vs. representative-ness; parametric vs. nonparametric options; “correlation is not causation”). Actively developing knowledge, skills and abilities required for statistical literacy.	Read and understand; reliably identify misspecification of methods chosen or employed. Choose and execute correct analysis, not necessarily able to choose the several methods that could be equally viable depending on investigator’s objectives. Qualified as a fluent, but not as an independent, statistical reasoner.	Understand scientific question to align statistical (or graphical) methods options to desired objectives. Expert review of technical features of proposal/paper-not necessarily of the science/statistics alignment. Qualified as independent experts in statistical reasoning.	Understand scientific question and clarify/encourage writer to clarify objectives so as to align statistical (or graphical) methods options to desired objectives. Expert review and evaluation—and diagnosis and remediation. Qualified to take individuals from pre-literate through to Master level statistical reasoning.
Not yet on the Bloom’s trajectory.	Bloom’s 1 remembering, understanding.	Bloom’s 2, 3, understand and apply but only apply what you’re told to apply.	Bloom’s 3–5 Choose and apply techniques. Analyze & interpret. Identify limitations, but not sophisticated enough to <i>independently</i> review literature, proposals, grants.	Bloom’s 5–6 evaluate (review) and synthesize for new methods but not for evaluation of others.	Bloom’s 6 synthesize for new methods, <i>and</i> evaluation of others.
Not a careful consumer.	Becoming a careful Consumer.		A careful consumer. Becoming a careful producer.	Expert consumer, expert producer.	
No or limited capacity to critique. Requires external “validation” to believe what is presented (e.g., “it was published in JAMA!” “Cochrane Reviews are correct”).		Developing; capacity to evaluate; sense of what is/is not appropriate; ability to critique; opinions on debates (e.g., application of multi-model inference; Bayesian vs. frequentist; when to use multiple-comparisons corrections).	Expert reviewer—capable of stewardship of the not-statistics discipline.	Expert review, diagnosis and recommender of remediation; capable steward of a statistical discipline.	

Table 1 includes an important column that does not actually appear in the Mastery Rubric but is included here because it is such a common stage across the biomedical and life sciences (and across some social and educational sciences as well): the pre-literate non-reasoner. This individual is described consistently in critiques of the quality of journal and grant reviewing (see also e.g., [27,29]), and is identified specifically by the lack of skills in the recent review by Weissgerber et al. (2016) [7]. The difference between a scientist who functions at this level and one who functions even at the Beginning Literacy level is *profound*—and their effects undermining the rigor and reproducibility of scientific research are increasingly less tolerable (e.g., [8–16]; see also [17]). Recognition that some reviews provided for journal editorial decisions, as well as grant funding, represent functioning at this level should be highlighted in these important decision-making contexts (i.e., even this structure represents actionable evidence).

The MR-SL can promote remediation of these identified weaknesses by individuals seeking to generate evidence they are “on the right track” or at least at the Beginning Literacy stage—and by institutions seeking to provide opportunities to achieve learning outcomes consistent with performance at this stage (or beyond). PhD students and scientists are often not operating at even the lowest Bloom’s taxonomy [53] level (knowledge—the main level targeted by most statistical training, see e.g., [21,27]) while professionally, they must function at the highest level (e.g., [10–12]; see also [32] (p. xiii)).

Evidence of this (perhaps surprisingly low) level of functioning with respect to statistical and quantitative argumentation comes from a variety of sources (e.g., [7,8,14–16]); the pre-literate non-reasoner is common and problematic. If evidence is found that an institution is training people to this level (and not beyond), action must be taken to remediate the situation or to reconfigure curriculum or learning objectives that purposefully aim at this level of performance. The MR-SL treats statistical literacy in a similar manner to general literacy [54]: comprising a set of learnable, improvable skills. In order to promote development of a CQP by initiating the learning and improving of this skillset, the MR-SL could be used to promote curricular or institutional remediation.

2.2. The Mastery Rubric for Statistical Literacy (MR-SL): KSAs for SL

The second dimension of a Mastery Rubric is the articulation of knowledge, skills, and abilities (KSAs) that are to be targeted and grown throughout the developmental trajectory. For the MR-SL, the list of KSAs representing statistical literacy was derived by synthesizing several models of statistical literacy with the more active “empirical enquiry” model of Wild and Pfannkuch ([57]; see also [58]). Because the developmental trajectory for these KSAs describes change from more naïve to more expert performance, the qualification of how these KSAs are executed is captured (and described) in the row that outlines growth and development in each KSA over time/training. The SL KSAs were synthesized from “A four-dimensional framework for statistical thinking in empirical enquiry” [52] (p. 19) and the “Statistical Thinking” facility described in [58] (p. 218) into the new MR-SL shown in Table 2.

Table 2. Mastery Rubric for Statistical Literacy (MR-SL).

Performance Level	Beginning Literacy	Functional Literacy	Skilled (Apprentice) Literacy	Independent (Journeyman) Literacy	Expert (Master)
General description of statistical literacy	Read, generally understand, notice gross errors, e.g., if categorical method applied to continuous variable or vice versa. Developing meta-cognitive awareness that if a question arises in their mind, the method may not be correct or clearly articulated. Engaging with statistical reasoning, developing awareness of this skill and how to grow/use it.	Consolidating reading and understanding, beginning to learn how to analyze (with software). Awareness of rules of thumb (e.g., sample size vs. representativeness; parametric vs. nonparametric options; “correlation is not causation”). Actively developing knowledge, skills and abilities required for statistical literacy.	Read & understand; reliably identify misspecification of methods chosen or employed. Choose and execute correct analysis, not necessarily able to choose the several methods that could be equally viable depending on research objectives. Qualified as a fluent, but not as an independent, statistical reasoner.	Understand scientific question to align statistical (or graphical) methods options to desired objectives. Expert review of technical features of proposal/paper-not necessarily of the science/statistics alignment. Qualified as independent expert in statistical reasoning.	Understand scientific question and clarify/encourage writer to clarify objectives so as to align statistical (or graphical) methods options to desired objectives. Expert review and evaluation—and diagnosis and remediation. Qualified to take individuals from pre-literate through to Master level statistical reasoning.
Considerations for evidence of performance at this level	Bloom’s 1 remembering, understanding.	Bloom’s 2, 3, understand and apply but only apply what you’re told to apply.	Bloom’s 3–5 <u>Choose and</u> apply techniques. Analyze and interpret. Identify limitations, but not sophisticated enough to <i>independently</i> review literature, proposals, grants.	Bloom’s 5–6 evaluate (review) and synthesize for new methods but not for evaluation of others.	Bloom’s 6 synthesize for new methods, <i>and</i> evaluation of others.
Define a problem based on critical literature review	Can identify the problem that is articulated within literature that is reviewed, but not derive or synthesize one across multiple sources. Does not question design features or evidence base supporting problems articulated in what was reviewed. Might argue that the impact factor of a journal as evidence that an article published there is “good” or “correct”.	Can identify the problem that is articulated within literature that is reviewed, and can recognize when incomplete review is provided. Does not derive or synthesize new issues from single or multiple sources. Acknowledges that design features and evidence base are essential for understanding the validity of claims or research problems articulated by others.	Can identify gaps and articulate problem (research questions) that arise from critical literature reviews, can recognize when incomplete review is provided and also recognizes the need to consider wider scope of literature for alternative solutions to a problem common across contexts or domains.	Can synthesize and define a theoretical or methodological problem based on a critical review of the literature in one or across scientific domains. Recognizes when and how solutions to problems from diverse contexts are or are not appropriate or adaptable for new applications.	Can diagnose and remediate individual synthesis and definitions of theoretical and/or methodological problems based on a critical review of the literature as well as critical evaluation of less expert synthesis <i>across contexts</i> —i.e., in terms of classroom work as well as grant proposals and manuscripts.
Identify or choose—and justify—the measurement properties of variables	Cannot identify the measurement system for variables within manuscripts unless they are articulated explicitly. If they are articulated, this information would not be useful/used.	Understands that there are different measurement systems but does not know how or why ratio-level data might be transformed into interval or ordinal data. Treats nominal data with numeric labels as if they are ratio-level.	Chooses measurement that optimizes power rather what specifically addresses hypothesis of interest. Limited consideration of interaction and mediation/moderation effects. Understands that nominal and ordinal data do not behave as ratio-level (or even integral) variables do.	Chooses measurement that optimizes generalizability and interpretability of results, and acknowledges that power may suffer—justifiably. Can justify (and recommend as appropriate) the transformation of data from one type to another if appropriate. Careful consideration of interaction and mediation/moderation effects.	Can identify and critique (as appropriate) the measurement system used in any given study/analysis. Can choose and justify nominal-, interval-, or ratio-level analytic methods. Understands the limitations of different types in terms of analysis assumption requirements, and can articulate the tradeoff in scientific explanatory power associated with measurement and data type choices. Expert consideration of interaction and mediation/moderation effects. Diagnosis and remediation of each of these <i>across contexts</i> .

Table 2. Cont.

Performance Level	Beginning Literacy	Functional Literacy	Skilled (Apprentice) Literacy	Independent (Journeyman) Literacy	Expert (Master)
Design the collection of data	Can identify data collection features in text if they match basic design elements from introductory materials (e.g., t-test, chi square) but cannot derive them if they are not present. Cannot design data collection initiatives. Cannot conceptualize covariates or their roles in analysis or interpretation.	Can identify data collection features if they are present in a manuscript/proposal—including more complex and advanced methods- but cannot derive them if they are not present. Recognizes covariates if mentioned, but does not require formal consideration (or justification) or evaluation of covariates.	Can match the correct data collection design to the instruments and outcomes of interest, but needs assistance in conceptualizing covariates and their potential roles in the planned analyses. May include covariates “because that is what is done” without being able to justify the roles of any in the hypotheses to be tested.	Can design appropriate data collection and identify instruments and outcomes (and covariates) that support the testing of specific hypotheses. Collaborates with expert as needed on appropriate use of advanced methods, including accommodating measurement and sampling error, attrition (if needed), and modeling requirements.	Expertly designs collection of data, including power calculations, modeling requirements, measurement/sampling error and data missingness. Designs and can critique sensitivity analyses as appropriate, and fluently diagnoses and remediates each of these <i>across contexts</i> .
Piloting, analysis and interpretation	Does not differentiate pilot studies and full studies; might not plan a pilot to ensure study features are feasible. Might call a study with a small N “pilot” just based on sample size. Cannot evaluate or interpret (their own or) others’ pilot work	Differentiates pilot and full-scale studies, but does not consider the ‘failures’ uncovered by pilot work to be informative-and might stop if pilot study uncovers problems. Might consider larger scale study unnecessary if pilot results are as expected.	Recognizes need for pilot studies and asks for appropriate assistance in the design and analysis. Pilot results are seen to be useful in addressing scalability issues. May seek assistance with scalability based on pilot results. Does not recognize when design or review demands are beyond their skill set.	Independently conceptualizes pilot studies that address relevant design issues. May seek expert advice for design, power, and analysis planning for their own work, and consistently recognizes when reviewing demands are beyond their skill set.	Expertly designs and analyzes pilot studies, utilizing the data for full study design, analysis planning and power, within their own and others’ work. Diagnoses and remediates each of these <i>across contexts</i> .
Discerning “exploratory”, “planned”, and “unplanned” data analysis	Does not perceive differences between “planned”, and “unplanned” data analysis in their own or others’ work. Does not recognize that exploratory analyses can be planned or unplanned and that these should be described as such.	In their own work, can differentiate between exploratory analysis and hypothesis testing, but not “planned” and “unplanned” analyses. May incorrectly characterize “exploratory” analysis as hypothesis testing (planned or unplanned).	Perceives differences between “planned”, and “unplanned” data analysis in their own work, but not in others’ work unless it is identified. May not recognize that exploratory analyses can be planned or unplanned, does not know why it might matter to communicate which they are doing/reporting.	Recognizes differences between “planned”, and “unplanned” data analysis in their own and others’ work, even when others do not recognize it in their own work. Knows that exploratory analyses can be planned or unplanned, and can identify which is included in their own and others’ work.	Clearly and consistently differentiates planned and unplanned analyses in their own work and that of others. Utilizes all types of analysis appropriately in support of coherent contributions to science. Consistently requires others to do the same, and can diagnose and remediate each of these <i>across contexts</i> in order to support scientific integrity and competence.
Hypothesis generation based on planned and unplanned analyses	Uses the default settings of software to guide analysis planning (and execution in the unplanned analysis case). Like software, does not differentiate planned or unplanned, nested or non-nested hypothesis tests. Does not generate hypotheses.	Uses the default settings of software to guide analysis planning (and execution in the unplanned analysis case). Attention is focused on planned analyses and hypothesis generation in that context; unlikely to generate testable hypotheses. May not recognize that hypotheses may be generated and tested in or by unplanned analyses or within the intermediate steps software executes to complete the desired analysis.	When software generates and tests hypotheses, treats that as “what was supposed to happen” and does not differentiate these results from those anticipated and resulting from planned analyses. Can generate new hypotheses, but is likely to base these on data without appeal to theory, plausibility, or context.	Can seamlessly integrate hypothesis generation into the consideration of literature or data analysis. In their own and others’ work, recognizes that, and articulates how, hypothesis generation from planned and unplanned analyses differ in their evidentiary weight and their need for independent replication. Depends on knowledge, context, and skills with synthesis—and not software—to generate testable hypotheses.	Expertly distinguishes hypothesis testing and hypothesis generation. Reliably recognizes and communicates the differences between these in all written and oral work. Consistently seeks to integrate plausibility and scientific contextualization into hypothesis generation. Diagnoses and remediates each of these <i>across contexts</i> .

Table 2. Cont.

Performance Level	Beginning Literacy	Functional Literacy	Skilled (Apprentice) Literacy	Independent (Journeyman) Literacy	Expert (Master)
Interpretation of results	Believes that the p -value is “true” and represents the evidence for the hypothesis or theory being tested. Never corrects for multiple comparisons in their own work; does not suggest or question the need for it in reviewing. Resists multiple comparisons corrections suggested by reviewers or collaborators if it causes “significant” results to disappear. Does not seek coherence in the analysis plan or the alignment of methods, results, and interpretation.	Understands that the p -value does not represent the “truth” of the hypothesis being tested, but cannot articulate why it is useful/used. Interprets p -values that are “very close” to the nominal alpha level (e.g., 0.049–0.10) as statistically meaningful evidence of trends; interprets very small p -values as “highly significant” results.	Understands that the p -value represents evidence supporting the null hypothesis, not the study hypothesis. Recognizes that very small p -values are not “highly significant results”, but does not consistently correct this language when reviewing. Can apply multiple comparisons corrections, but does so when reminded. Does not insist on these corrections in work that they review (grants, manuscripts, coursework).	Understands that the null hypotheses that statistical tests test are <i>never</i> the actual purpose of the analysis. Resists reification and is committed to good-faith efforts to falsify hypotheses, not simply test the null. Applies multiple comparisons to promote reproducible results. In their own and others’ work, seeks competing, plausible, alternative models or explanations.	Communicates consistently that the null hypotheses that statistical tests test are <i>never</i> the actual purpose of the analysis. Resists reification and is committed to good-faith efforts to falsify hypotheses, not simply test the null. Seeks competing, plausible, alternative models or explanations. Encourages collaborators to do all of these, and diagnoses and remediates each of these <i>across contexts</i> .
Draw and contextualize conclusions	p -value driven conclusions without consideration of limitations. No contextualization of the results with prior literature or with the foregoing portions of the document. Conclusions may not actually represent results; overinterpretation and failure to identify or acknowledge limitations.	p -value—driven conclusions that may include consideration of limitations including multiple comparisons. Conclusions are typically superficial—i.e., not very deeply contextualized with the literature. Conclusions are typically aligned with results, but may not be well-contextualized with the rest of the document (paper, grant).	In their own work, draws conclusions that are contextualized with the entire manuscript/grant. In reviewing, does not require that conclusions are aligned with the whole document, and does not require full contextualization. Incomplete consideration of limitations in their own work and inconsistent requirement that limitations are acknowledged in others’ work.	Contextualizes results with respect to the entirety of the manuscript/grant, and so can detect cases where conclusions are not aligned with the introduction/background, methods, and/or results. Careful consideration of limitations deriving from the method and its application in the specific study. Requires full contextualization of conclusions in others’ work and strives to fully contextualize conclusions in their own work.	Expertly differentiates effect sizes, clinical significance and statistical significance. Can articulate either multi-trait/multi-method (MMTM) or other triangulation approach, including mixed methods analysis to understand and contextualize results. Consistently requires full contextualization of conclusions in others’ work and fully contextualizes conclusions in their own work. Diagnoses and remediates each of these <i>across contexts</i> .
Communication	Does not communicate statistical information clearly or consistently, skips the methods section of papers or grants. Does not differentiate appropriate and inappropriate communication with statistics or other quantitative material. Does not generate or evaluate communication of statistical or quantitative material.	Reads the statistics and methods sections superficially. Does not recognize inconsistencies (e.g., author describes data as categorical and plans t-test). May state that “only the p -value is needed” when reviewing how results are communicated. Does not generate communication of statistical or quantitative material and should not review these.	Reads the statistics and methods sections and identifies what they are and are not able to review competently. Can formulate queries for either the author or for an expert to help them complete a review. Seeks to collaborate with statistical expert to ensure that team-based reporting is coherent, consistent, and accurate.	Consistent proficient use of statistical and quantitative language to correctly describe what was done, why, and how. Sufficient consideration given to limitations with explicit contextualization of results consistently included in the interpretation of results. Errors of comprehension of this text—if they arise—arise on the side of the reader.	Expert communicator and reviewer of scientific communication relating to or including statistical and quantitative materials. Consistent sensitivity to audience and appropriate interpretation and contextualization of results. In reviewing proposals, can anticipate (diagnose) challenges for dissemination and communication, and differentiate errors in reasoning from failures to disclose or articulate. Diagnoses and remediates each of these <i>across contexts</i> .

The model of statistical thinking articulated by Wild and Pfannkuch ([57] discussed in [52] (pp. 18–20), captures the features of literacy, reasoning, and thinking that are relevant for graduate science curricula (as noted by [58]; see also [50]) and beyond. Thus, this model embodies “ . . . value on the integration of quantitative methods as part of the substantive enterprise of doctoral education” [12] (p. 236). The KSAs (rows) in the MR-SL are:

- Define a problem based on critical literature review;
- Identify or choose—and justify—the measurement system;
- Design the collection of data;
- Piloting, analysis and interpretation;
- Discerning “exploratory”, “planned”, and “unplanned” data analysis;
- Hypothesis generation based on planned and unplanned analyses;
- Interpretation of results;
- Draw and contextualize conclusions;
- Communication.

These KSAs generally define the scientific method—and also require content knowledge. The initiation and development of these KSAs could therefore be integrated across multiple content course areas, and also for those who are practicing scientists—whether or not they completed PhD-level training. The MR-SL serves to link instruction in statistical methods with the application, and reasoning with, those methods and results. Thus, it can support the initiation of the development of this set of KSAs and their continued promotion within, and beyond the ending of, formal education. There are six mutually exclusive performance level descriptors for each of these KSAs in the MR-SL (Table 2); the integration of the Bloom’s level functioning at different stages with the features of statistical literacy are explicit.

3. Results

The MR-SL in Table 2 co-articulates cognitive perspectives on the development (columns) derived from extant literature with context-appropriate and explicit, but flexible, descriptions of a complex set of knowledge skills and abilities KSAs (rows) that represent statistical literacy as a learnable, improvable skill set. Table 3 provides a rough alignment of the KSAs in the MR-SL and definitions of statistical “literacy”, “thinking”, and “reasoning” in prior models.

Table 3. Alignment of models of statistical reasoning, thinking, and literacy with the MR-SL KSAs.

Tractenberg MR-SL KSAs (defining Statistical Literacy Like Chall [55] Defined General Literacy: As a Learnable and Improvable Skill Set)	Bishop and Talbot 2001 [58] (statistical Thinking)	Wild and Pfannkuch 1999 [57] (statistical Thinking)	Garfield, delMas, Chance 2003 [34] (Definitions of Statistical Literacy, Thinking, Reasoning)
Define a problem based on critical literature review.	Identify the problem.	Constructing and reasoning from models.	Statistical thinking.
Identify or choose—and justify—the measurement properties of variables.	Plan the experiment/survey/observational study.	Taking account of variation; Constructing and reasoning from models.	Statistical thinking.
Design the collection of data. Piloting, analysis and interpretation.	Pilot and adjust (analyze and interpret the data)		
Discerning “exploratory”, “planned”, and “unplanned” data analysis.	Do final study; collect and present the data; analyze and interpret the data.	Constructing and reasoning from models; transnumeration (transforming data for understanding);	
Hypothesis generation based on planned and unplanned analyses.		synthesis of problem context and statistical understanding.	
Interpretation of results.	“To think statistically means that one can:		Statistical reasoning.
Draw and contextualize conclusions.	1. Read data, critically and with comprehension;		
Communicate.	2. Produce data that provide clear answers to important questions;		
	3. Draw trustworthy conclusions based on data” [58] (p. 220).		

Table 3 shows that the alignment of the MR-SL KSAs is tightest with the model of statistical thinking outlined by Bishop and Talbot (2001) [58] but it is also similar to the statistical thinking model of Wild and Pfannkuch (1999) [57]. The definition of “statistical literacy” given in Garfield, delMas and Chance (2003) [34] may be foundational to engaging in any of the KSAs, but “statistical thinking” as they defined it may be more aligned with the *consumption* of statistical reasoning and is not (according to Table 3) involved in production. However, considering the alignment of their definition of “statistical reasoning” with the interpretation of results and drawing of conclusions suggests that the ability to reason statistically can be developed without a focus on data collection or analysis (which are key aspects of the other two models of “statistical thinking”). For scientists who are either in training or in practice, both the production and consumption of statistical argumentation are essential and these can be leveraged as two types of important practice for ensuring that the learning in statistics coursework is sustainable (endures beyond the end of teaching and can be applied in different contexts than where it was learned).

It is not necessary that all learners progress on all KSAs simultaneously; with the co-articulation of KSAs with developmental stages of performance on each one, instructors, institutions and students can leverage their time and effort in order to ensure that all KSAs are performed, at some point (e.g., midway through a degree program), at a target level. The co-articulation of the MR-SL also supports the generation of actionable evidence for learners who identify one or another KSA as most challenging, as well as for institutions or instructors that identify performance on one KSA or another as least-consistent across a student cohort. The co-articulation also both captures an explanation for why reproducibility and peer review in science are widely perceived to be weak (i.e., because people do “function” at insufficiently-sophisticated levels on some, if not all, of these important KSAs) and also provides an approach, to an individual, instructor or the institution, for addressing this weakness.

Table 4 is a degrees of freedom analysis [47–49] evaluating the alignment of the MR-SL and its potential to support evidence-based decision-making in higher education as well as the five Principles for the documentation of learning outcomes [46].

It can be seen in Table 4 that four of five Principles [46] are addressed by the creation of the MR-SL and its adoption to promote statistical literacy that is appropriate for PhD science students and anyone who will consume or produce scientific argumentation that depends on data or quantitation. One of the five principles (collaborative) is not addressed at all by the MR-SL; however, the MR-SL KSAs are articulated based on multiple models of statistical thinking and reasoning, and on real-world requirements for applied statistical literacy by scientists in their daily work. Explicit articulation allows learners to see what is expected of them and institutions/instructors to support learners in their achievement of these expectations. Thus, the MR-SL is “representative”, but not necessarily “collaborative”. Its implementation at any institution would need to be based on all stakeholders agreeing, so “collaboration” might become relevant in that (implementation) context.

The alignment of the MR-SL with one of the principles (“Outcomes are focused on *improvement*”) does not differentiate between the learner and the instructor/institution. The instructor/institution can obtain actionable evidence about how courses or training support improvement in key outcomes, and the learner can obtain actionable evidence about what other information or training is needed in order to achieve a targeted performance level on each KSA. With a Mastery Rubric self-monitoring is focused on, “what training do we offer to promote growth or development of this KSA, what else can we offer to support if for all learners?” (institution/instructor) and “how well do I do/know this KSA, what do I need to do to become more proficient?” (learner). These perspectives are sufficiently similar to warrant collapsing over instructor/learner in Table 4.

Table 4. Alignment of Principles for documenting and improving assessment with features of the MR-SL from student and institutional perspectives.

Principles for Documenting/Improving	Student Performance	Institutional Effectiveness
Develop/articulate specific actionable learning outcomes	MR-SL helps students identify their progress towards articulated learning objectives.	MR-SL helps instructors/institutions identify and articulate developmental learning objectives.
Connect learning goals with student work	If work is not explicitly aligned with learning goals, students see this and can remediate that (with additional work or training).	If learning goals are not reflected in student work (assignments), instructors/institution can see this and remediate with different assignments.
Articulate learning outcomes collaboratively	<i>Not addressed by the MR-SL.</i>	
Outcomes support assessment that generates actionable evidence	Students can/are encouraged to actively self assess, to ensure they are making progress on the developmental path.	Institutions and instructors see explicit alignment of curricular features (courses, assignments/work products) and can use this evidence to support or change the approach.
Outcomes are focused on <i>improvement</i>	The explicit articulation of expected growth and development in the target KSAs focuses all stakeholders on improvement of these KSAs—emphasizing they are not static.	
Outcomes document learning and its extent	Learners generate evidence of their achievement and ongoing development of KSAs.	Instructors/institutions structure training/teaching to generate documentation of learning and the achievement of articulated learning objectives.
Outcomes provide evidence of quality of learning	A portfolio can be created articulating the extent and quality of learning.	Assessment opportunities that document the achievement and quality of learning can be developed.
Expectations are explicit in the outcomes	The MR-SL makes explicit the expectation that the learner takes some responsibility for self-assessment and ensuring ongoing development until the target performance level is achieved.	The MR-SL makes explicit the institutional obligation to provide learning opportunities that can and do promote growth and development in the target KSAs.
Evidence from the outcomes is externally relevant	Portfolios documenting the achievement of learning outcomes (in statistical literacy) can be used to document readiness/qualification to review.	Statistical literacy is known to be weak; institutions that adopt the MR-SL and use it to guide curriculum development or evaluation can document their alignment of learning outcomes with the improvement of statistical literacy and/or contributions to the collective quantitative proficiency.

Four additional rows are included in Table 4. These are not “principles” for documenting learning per se, but they are relevant to a discussion about promoting statistical literacy with a Mastery Rubric approach and they are also discussed in the NILOA policy statement [46]. These additional considerations are that:

- Outcomes document learning and its extent;
- Outcomes provide evidence of quality of learning;
- Expectations are explicit in the outcomes; and
- Evidence from the learning outcomes is externally relevant.

4. Discussion

A Mastery Rubric emphasizes habits of mind as they transition from more novice to more expert, along a Bloom’s-compatible developmental trajectory [37]. The developmental stages of the MR-SL map onto those articulated for the development of general literacy [55], and the potential for explicit articulation of performance levels for each KSA are aligned with the self-efficacy argument of Bandura [59] (Chapter 2). The MR-SL captures key features of engagement in scientific inquiry (e.g., [57,58]); it is consistent with four of five NILOA principles for learning outcomes, and has an additional four features that generate actionable evidence by both the learner and the instructor/institution. Overall, these features suggest that the MR-SL is strongly supportive of the perspective that documenting learning matters.

Undergraduate statistical literacy is fundamentally different from that required for applied science and for doctoral level work, but it is not expertise *in statistics* that is targeted with the

MR-SL—it is expertise, or movement towards it, in this particular type of *literacy* that is targeted. Perhaps especially, explicitly describing what the KSAs are and how they should be developing can promote the recognition that/when additional training (institution) or information (learner) is needed. This approach is supportive of the identification of strengths and weaknesses—in the student and in the curriculum—thereby promoting creation or revision of training opportunities to address the identified gaps. This in turn can promote the concrete articulation by learners/trainees of how statistical training experiences have actually promoted observable changes in their own SL strengths and weaknesses.

The one NILOA principle for documenting learning outcomes with which the MR-SL is not closely aligned, that outcomes are articulated via collaboration with all stakeholders, is a significant limitation to the applicability of this work—because buy-in from faculty across courses and possibly disciplines or departments is critical for institutional adoption of the MR-SL. It is possible, however, that a focus on the representativeness of the KSAs of what is commonly defined as “statistical literacy”, and the alignment of the developmental trajectory with other well-established models (e.g., [54,57,58]; [59] (Chapter 2)), can facilitate consideration of how the MR-SL can best be adopted or adapted to achieve institutional objectives in support of the collective quantitative proficiency and statistical literacy that modern scientific practice requires.

5. Conclusions

Many PhD science programs require a single statistics course, and although this may suffice for undergraduates (see [31]), statistical literacy to support responsible stewardship of a scientific discipline differs fundamentally from that of undergraduates (see, e.g., [58]). Any syllabus can be compared to the MR-SL to determine the stage at which learners would be able/are expected to function, as well as evaluating the extent to which the learning objectives articulated in the syllabus are supportive of growth in statistical literacy. Moreover, the MR-SL can be used like other Mastery Rubrics have been to revise an existing curriculum [43], or to create new training opportunities that can promote the initiation of, and sustainable development in, a target set of KSAs [42].

It is not possible for degree and training programs to teach every quantitative method. At a minimum, because the developmental trajectory and KSAs are specified in this article, institutions can use it to determine the highest level a graduate in any program (undergraduate or graduate) can expect to attain given the existing statistical training and practice opportunities. Individual scientists may use the MR-SL to seek new quantitative learning opportunities by placing themselves on the developmental trajectory with respect to each KSA. With a focus on their metacognitive awareness of their own statistical literacy, individuals in or beyond their formal education setting can discern their growth and/or the need for more training.

The National Institute for Learning Outcomes Assessment [46] articulates that . . . “students need a postsecondary education that will prepare them to meet the challenges of the 21st century” (see also [60]). This paper describes a model for statistical literacy (SL) and its development that can support the dynamics of practicing modern science, starting with either graduate or undergraduate training. The MR-SL does so by generating actionable evidence about learning outcomes in statistical literacy from institutional, instructor, and student performance. The developmental framework around SL promotes the learner’s understanding of his/her own statistical reasoning, as well as growth and depth of their knowledge, skills, and abilities relating to data and statistical analysis. This feature is an important element of education that can prepare learners “to meet the challenges of the 21st century”, because knowledge is increasing at a rate we simply cannot keep up with. A crucial aspect of 21st century education is preparing individuals to *continue learning*—part of which involves self-assessment.

The Mastery Rubric provides explicit opportunities for consequential assessment that serves students, instructors, developers/reviewers/accreditors of a curriculum, and institutions. By supporting the enrichment, rather than increasing the amount, of statistical training in the sciences, the MR-SL supports evaluable curriculum development, evaluation, and delivery to promote statistical literacy for students and a collective quantitative proficiency more broadly. This model for promoting

SL can be adopted by an individual for their own learning, or by a department or discipline, to promote ongoing and integrated teaching and learning in statistical reasoning. The extent that the model is adopted can support a cultural shift across scientific disciplines towards a collective quantitative proficiency that enables scientists and students alike to determine which methods to learn about and also how to know if they have learned enough about the chosen methods for professional-level engagement in modern life and science.

Acknowledgments: This work was completed without grant funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Steen, L.A. (Ed.) *Mathematics and Democracy: The Case for Quantitative Literacy*; The Woodrow Wilson Fellowship Foundation: Princeton, NJ, USA, 2001.
2. Watson, J.M.; Callingham, R. Statistical literacy: A complex hierarchical construct. *Stat. Educ. Res. J.* **2003**, *2*, 3–46.
3. Carvalho, C.; Solomon, Y. Supporting statistical literacy: What do culturally relevant/realistic tasks show us about the nature of pupil engagement with statistics? *Int. J. Educ. Res.* **2012**, *55*, 57–65. [[CrossRef](#)]
4. Schuyten, G. Discussion: Research skills: A closely connected triplet of research area, research methodology, and statistics. In *Training Researchers in The use of Statistics: IASE Round Table Conference*; Batanero, C., Ed.; International Association for Statistical Education, International Statistical Institute: Voorburg, The Netherlands, 2001; pp. 227–230.
5. Steen, L.A. *Achieving Quantitative Literacy: An Urgent Challenge for Higher Education*; MAA notes; Mathematical Association of America: Washington, DC, USA, 2004; Volume 62.
6. Nikiforidou, Z.; Leek, A.; Pange, J. Statistical literacy at university level: The current trends. *Procedia Soc. Behav. Sci.* **2010**, *9*, 795–799. [[CrossRef](#)]
7. Weissgerber, T.L.; Garovic, V.D.; Milin-Lazovic, J.S.; Winham, S.J.; Obradovic, Z.; Trzeciakowski, J.P.; Millic, N.M. Reinventing biostatistics education for basic scientists. *PLoS Biol.* **2016**, *4*, e1002430. [[CrossRef](#)] [[PubMed](#)]
8. Collins, F.S.; Tabak, L.A. Policy: NIH Plans to enhance reproducibility. *Nature* **2014**, *505*, 612–613. [[CrossRef](#)] [[PubMed](#)]
9. Cumming, G.; Fidler, F.; Vaux, D.L. Error bars in experimental biology. *J. Cell Biol.* **2007**, *177*, 7–11. [[CrossRef](#)] [[PubMed](#)]
10. Finch, S.; Cumming, G.; Thomason, N. Reporting of statistical inference in the Journal of Applied Psychology: Little evidence of reform. *Educ. Psychol. Meas.* **2001**, *61*, 181–210. [[CrossRef](#)]
11. Fidler, F.; Thomason, N.; Cumming, G.; Finch, S.; Leeman, J. Editors can lead researchers to confidence intervals, but can't make them think. *Psychol. Sci.* **2004**, *15*, 119–126. [[CrossRef](#)] [[PubMed](#)]
12. Henson, R.K.; Hull, D.M.; Williams, C.S. Methodology in our education research culture: Toward a stronger collective quantitative proficiency. *Educ. Res.* **2010**, *39*, 229–240. [[CrossRef](#)]
13. Hayden, E.C. Weak statistical standards implicated in scientific irreproducibility. *Nat. News* **2013**. [[CrossRef](#)]
14. Ioannidis, J.P.A. Why most published research findings are false. *PLoS Med.* **2005**, *2*, e124. [[CrossRef](#)] [[PubMed](#)]
15. Vaux, D.L. Research methods: Know when your numbers are significant. *Nature* **2012**, *492*, 180–181. [[PubMed](#)]
16. Vaux, D.L. Basic statistics in cell biology. *Annu. Rev. Cell. Dev. Biol.* **2014**, *30*, 23–37. [[CrossRef](#)] [[PubMed](#)]
17. Baker, M. Is there a reproducibility crisis? *Nature* **2016**, *533*, 452–454. [[CrossRef](#)] [[PubMed](#)]
18. Golde, C.; Walker, G. (Eds.) *Envisioning the Future of Doctoral Education: Preparing Stewards of the Discipline—Carnegie Essays on the Doctorate*; Jossey-Bass: San Francisco, CA, USA, 2006.
19. Shulman, L.S. Foreward. In *The Formation of Scholars: Rethinking Doctoral Education for the Twenty First Century*; Walker, G.E., Golde, C.M., Jones, L., Bueschel, A.C., Hutchings, P., Eds.; Jossey-Bass: San Francisco, CA, USA, 2008; pp. 9–13.

20. Aiken, L.S.; West, S.G.; Millsap, R.E. Doctoral training in statistics, methodology, and measurement in psychology: Replication and extension of Aiken, West, Sechrest and Reno's (1990) survey of PhD programs in North America. *Am. Psychol.* **2008**, *63*, 32–50. [[CrossRef](#)] [[PubMed](#)]
21. Aiken, L.S.; West, S.G.; Millsap, R.E. Improving training in methodology enriches the science of psychology. *Am. Psychol.* **2009**, *64*, 51–52. [[CrossRef](#)] [[PubMed](#)]
22. Zimiles, H. Ramifications of increased training in quantitative methodology. *Am. Psychol.* **2009**, *64*, 51. [[CrossRef](#)] [[PubMed](#)]
23. Batanero, C. (Ed.) *Training Researchers in the Use of Statistics: IASE Round Table Conference*; International Association for Statistical Education; International Statistical Institute: Voorburg, The Netherlands, 2001.
24. Van Horn, J.D. Opinion: Big data biomedicine offers big higher education opportunities. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 6322–6324. [[CrossRef](#)] [[PubMed](#)]
25. Welch, L.; Lewitter, F.; Schwartz, R.; Brooksbank, C.; Radivojac, P.; Gaeta, B.; Schneider, M.V. Bioinformatics curriculum guidelines: Toward a definition of core competencies. *PLoS Comput. Biol.* **2014**, *10*, e1003496. [[CrossRef](#)] [[PubMed](#)]
26. American Association for the Advancement of Science. Vision and Change in Undergraduate Biology Education: A Call to Action. 2009. Available online: <http://www.visionandchange.org/> (accessed on 7 July 2011).
27. Hellems, M.A.; Gurka, M.J.; Hayden, G.F. Statistical literacy for readers of Pediatrics: A moving target. *Pediatrics* **2007**, *119*, 1083–1088. [[CrossRef](#)] [[PubMed](#)]
28. Windish, D.M.; Huot, S.J.; Green, M.L. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA* **2007**, *298*, 1010–1022. [[CrossRef](#)] [[PubMed](#)]
29. Anderson, B.; Williams, S.; Schulkin, J. Statistical literacy of obstetrics-gynecology residents. *J. Grad. Med. Educ.* **2013**, *5*, 272–275. [[CrossRef](#)] [[PubMed](#)]
30. Garfield, J.; Hogg, B.; Schau, C.; Whittinghill, D. First courses in statistical science: The status of educational reform efforts. *J. Stat. Educ.* **2002**, *10*. Available online: <https://ww2.amstat.org/publications/jse/v10n2/garfield.html> (accessed on 7 July 2011).
31. American Statistical Association (ASA). *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*; ASA: Alexandria, VA, USA, 2005.
32. Hancock, G.R.; Mueller, R.O. (Eds.) *The Reviewer's Guide to Quantitative Methods in the Social Sciences*; Routledge: New York, NY, USA, 2010.
33. Tractenberg, R.E. Integrating ethical reasoning into preparation to work in/with Big Data through the Stewardship model. In *Ethical Reasoning in Big Data: An Exploratory Analysis*; Collmann, J., Matei, S., Eds.; Springer: New York, NY, USA, 2016; pp. 185–192.
34. Garfield, J.; delMas, R.; Chance, B. The Web-Based ARTIST: Assessment Resource Tools for Improving Statistical Thinking. In American Educational Research Association Meeting, Chicago, IL, USA, 21–25 April 2003. Available online: https://app.gen.umn.edu/artist/articles/AERA_2003.pdf (accessed on 2 May 2010).
35. Hutchings, P.; Kinzie, J.; Kuh, G.D. Evidence of student learning: What counts and what matters for improvement. In *Using Evidence of Student Learning to Improve Higher Education*; Kuh, G.D., Ikenberry, S.O., Jankowsk, N.A., Eds.; Jossey-Bass: Somerset, NJ, USA, 2015; pp. 27–50.
36. Stevens, D.D.; Levi, A.J. *Introduction to Rubrics: An Assessment Tool to Save Grading Time, Convey Effective Feedback and Promote Student Learning*; Stylus Publishing: Portland, OR, USA, 2005.
37. Tractenberg, R.E.; McCarter, R.J.; Umans, J. A mastery rubric for clinical research training: Guiding curriculum design, admissions, and development of course objectives. *Assess. Eval. High. Educ.* **2010**, *35*, 15–32. [[CrossRef](#)] [[PubMed](#)]
38. Tractenberg, R.E. Developing a curriculum for research in pathology residency: A Pathology Research Mastery Rubric. Poster presented at the Northeast Group on Educational Affairs (NEGEA) Annual Meeting, Washington, DC, USA, 10 October 2011.
39. Zalles, D.; Haertel, G.; Mislevy, R.J. *Using Evidence-Centered Design to Support Assessment, Design and Validation of Learning Progressions*; Large Scale Assessment Technical Report 10; SRI International: Menlo Park, CA, USA, 2010. Available online: http://ecd.sri.com/downloads/ECD_TR10_Learning_Progressions.pdf (accessed on 22 July 2011).

40. Lustbader, P. Construction Sites, Building Types, and Bridging Gaps: A Cognitive Theory of the Learning Progression of Law Students. *Willamette L. Rev.* **1997**, *33*, 315. Available online: http://heinonline.org/HOL/Page?handle=hein.journals/willr33&div=16&g_sent=1&collection=journals (accessed on 15 June 2011).
41. Schwarz, C.V.; Reiser, B.J.; Davis, E.A.; Kenyon, L.; Achér, A.; Fortus, D.; Shwartz, Y.; Hug, B.; Krajcik, J. Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *J. Res. Sci. Teach.* **2009**, *46*, 632–654. [[CrossRef](#)]
42. Tractenberg, R.E.; FitzGerald, K.T. A mastery rubric for the design and evaluation of an institutional curriculum in the responsible conduct of research. *Assess. Eval. High. Educ.* **2012**, *37*, 1003–1021. [[CrossRef](#)]
43. Tractenberg, R.E.; Gushta, M.M.; Weinfeld, J. The mastery rubric for evidence-based medicine: Institutional validation via multi-dimensional scaling. *Teach. Learn. Med.* **2016**, *28*, 152–165. [[CrossRef](#)] [[PubMed](#)]
44. Messick, S. The interplay of evidence and consequences in the validation of performance assessments. *Educ. Res.* **1994**, *23*, 13–23. [[CrossRef](#)]
45. Toohey, S. *Designing Courses for Higher Education*; The Society for Research into Higher Education & Open University Press: Philadelphia, PA, USA, 1999.
46. National Institute for Learning Outcomes Assessment. *Higher Education Quality: Why Documenting Learning Matters*; University of Illinois and Indiana University: Urbana, IL, USA, 2016.
47. Campbell, D.T. “Degrees of freedom” and the case study. *Comp. Political Stud.* **1975**, *8*, 178–193.
48. Woodside, A.G. *Case Study Research: Theory, Methods and Practice: Theory, Methods, Practice*; Emerald Group: Bangles, UK, 2010.
49. Tractenberg, R.E. Degrees of freedom analysis in educational research: Ensuring the capstone project functions as assessment. **2016**, submitted.
50. Jones, G.A.; Langrall, C.W.; Mooney, E.S.; Thornton, C.A. Models of development in statistical reasoning. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Ben-Zvi, D., Garfield, J., Eds.; Springer: Dordrecht, The Netherlands, 2004.
51. Ben-Zvi, D.; Garfield, J. Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. In *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*; Ben-Zvi, D., Garfield, J., Eds.; Kluwer Academic Publishers: Amsterdam, The Netherlands, 2004; pp. 3–15.
52. Pfannkuch, M.; Wild, C.J. Towards an understanding of statistical thinking. In *The challenge of Developing Statistical Literacy, Reasoning and Thinking*; Ben-Zvi, D., Garfield, J., Eds.; Kluwer Academic Publishers: Amsterdam, The Netherlands, 2004; pp. 17–46.
53. Bloom, B.S.; Engelhart, M.D.; Furst, E.J.; Hill, W.H.; Krathwohl, D.R. *Taxonomy of Educational Objectives: Handbook I: Cognitive Domain*; David McKay: New York, NY, USA, 1956.
54. Anderson, L.W.; Krathwohl, D.R.; Airasian, P.W.; Cruikshank, K.A.; Mayer, R.E.; Pintrich, P.R.; Raths, J.; Wittrock, M.C. (Eds.) *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom’s Taxonomy of Educational Objectives*; Addison Wesley Longman, Inc.: White Plains, NY, USA, 2001.
55. Chall, J.S. *Stages of Reading Development*; Five Stages of Reading Development; McGraw-Hill Book Company: New York, NY, USA, 1983. Available online: http://tools4reading.com/web/wp-content/uploads/2015/05/challs_stages_of_reading_development.pdf (accessed on 1 March 2010).
56. Garfield, J.; delMas, R.; Zieffler, A. Assessing important learning outcomes in introductory tertiary statistics courses. In *Assessment Methods in Statistical Education*; Bidgood, P., Hunt, N., Joliffe, F., Eds.; John Wiley & Sons: New York, NY, USA, 2010; pp. 75–86.
57. Wild, C.J.; Pfannkuch, M. Statistical thinking in empirical enquiry. *Int. Stat. Rev.* **1999**, *67*, 223–265. [[CrossRef](#)]
58. Bishop, G.; Talbot, M. Statistical thinking for novice researchers in the biological sciences. In *Training Researchers in the Use of Statistics: IASE Round Table Conference*; Batanero, C., Ed.; International Association for Statistical Education, International Statistical Institute: Voorburg, The Netherlands, 2001; pp. 215–226.
59. Bandura, A. *Self-Efficacy: The Exercise of Control*; WH Freeman: New York, NY, USA, 1997.
60. Kuh, G.D.; Ikenberry, S.O.; Jankowski, N.; Cain, T.R.; Ewell, P.T.; Hutchings, P.; Kinzie, J. *Using Evidence of Student Learning to Improve Higher Education*; Jossey-Bass: San Francisco, CA, USA, 2015.

