Rubric Authoring Tool Supporting Cognitive Skills Assessment across an Institution

Simper, Natalie

Natalie Simper, QUEEN'S UNIVERSITY, natalie.simper@queensu.ca

Rubric Authoring Tool Supporting Cognitive Skills Assessment Across an Institution

ABSTRACT

This paper explores a method to support instructors in assessing cognitive skills in their course, designed to enable aggregation of data across an institution. A rubric authoring tool, "BASICS" (Building Assessment Scaffolds for Intellectual Cognitive Skills) was built as part of the Queen's University Learning Outcomes Assessment (LOA) Project. It provides a workflow for assessment choices and generates an assessment rubric that can be tailored to individual needs based on user input. The dimensions and criteria in BASICS were adapted from the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics, and drew on annotations from over 900 work samples from the LOA project. This paper summarizes the development of the tool, and presents initial reliability and validity data from a pilot study. The pilot found that the BASICS developed rubric was consistent for the assessment of critical thinking and problem solving. The pilot compared assessment data derived from course Teaching Assistants with that of trained Research Assistants. Analysis found moderate intraclass correlation coefficients between the BASICS rubric and corresponding VALUE rubric dimensions, suggesting that the BASICS rubric aligned with the VALUE criteria. Preliminary findings suggest that BASICS is an effective tool for instructors to author rubrics, tailored to their own specifications for assessment of cognitive skills in a course. It is also promising as a method for aggregation of data across the institution. Researchers are conducting further investigation to evaluate the reliability of BASICS rubrics over multiple work samples from a range of disciplinary contexts.

KEYWORDS

Assessment, rubric, critical thinking, creative thinking, problem solving

INTRODUCTION

Many consider assessment to be the most significant prompt for learning (Boud, 1995; Brown, 2004) and assessment cultures are moving toward assessing higher order thinking processes and competences instead of factual knowledge and lower level cognitive skills (Jonsson & Svingby, 2007). In addition, there has been increasing pressure from funding agencies for higher education institutions to provide evidence of skill development across the institution (Herbert, 2015). To provide this evidence, however, often requires change in assessment practices. The challenges to encouraging change in assessment practices include navigating status quo (assessment culture), the existing knowledge of assessment principles and the capacity of instructors to incorporate reliable, evidence-based practices.

Reliability of assessment in a course is important so that consistent judgements of student performance are made. For institutional purposes, such as informing program improvement or for accountability (Kuh, Jankowski, Ikenberry, & Kinzie, 2014), assessment needs to be reliable, not only within the course, but consistent enough to allow data to be aggregated across courses. Assessment rubrics are one method for clarifying outcomes and communicating expectations to students, and are

commonly used to evaluate the quality of students' responses. By using rubrics, instructors can effectively communicate what they intended students to demonstrate (Wiggins & McTighe, 2005).

The American Association of Colleges and Universities (AAC&U) designed the Valid Assessment of Learning in Undergraduate Education (VALUE) rubrics as part of the Liberal Education and America's Promise (LEAP) initiative, for broad use in higher education as guides for assessment of essential learning outcomes (Partnership for 21st Century Skills, 2008). There are sixteen rubrics in total, each comprising four levels of performance criteria that range from a low benchmark level to a high capstone performance level. "As of December 2015, the rubrics have been accessed by more than 42,000 individuals from more than 4,200 unique institutions, including more than 2,800 colleges and universities" (VALUE, 2014). However, using the VALUE rubrics is time and labour intensive and usually requires that marking is conducted as an activity separate from course grading (Siefert, 2011).

Many institutions utilize a VALUE rubric assessment program as part of professional development provided for faculty. Some institutions provide financial incentives such as stipends to encourage faculty participation, and often include workshop series run by institutional teaching centres (Bernstein & Greenhoot, 2014). As beneficial as these professional development opportunities are for instructors, they seldom link directly to course assessment. Another suggestion for using the VALUE rubrics is to adapt them for local contexts (T. L. Rhodes, 2011); this is sometimes done as part of larger program improvement initiatives (Gray, 2013; Clark & Eynon, 2011). Adaptation of the rubrics for use at the program or course level requires expertise, input from stakeholders, and agreement among instructors. Even with these aspects in place, the resulting rubrics do not specifically describe the outcomes at the assignment level.

The Learning Outcomes Project (LOA) (Frank, Simper, & Kaupp, 2016) used three of the VALUE rubrics (T. L. Rhodes & Finley, 2013), applied to a selection of work artifacts (assignments) with the goal of evaluating the growth of cognitive skills within a program and across learning areas. The VALUE rubrics were used in a relatively controlled manner; a project coordinator facilitated training of Research Assistants, with calibration sessions and cross marking (see Method section for details). The LOA project was a four-year longitudinal study supported by funding from the Higher Education Quality Council of Ontario (HEQCO). Long-term sustainability was a goal of the project, requiring a mechanism to move the work beyond the bounds (and funding) of research and into ongoing practices at the institution. Rather than train instructors in the use of the very general VALUE rubrics, researchers developed a rubric authoring tool called "BASICS" (Building Assessment Scaffolds for Intellectual Cognitive Skills) for the adaptation of the critical thinking, creative thinking, and problem solving rubrics at the course assignment level. Presented here is the development and pilot of the tool. The overall results from the LOA project will be published through the funding agency (HEQCO).

BASICS was designed to target a four-part problem; namely,

1. The need to aggregate across the institution,
2. Instructors' need for rubrics that can be used at the assignment level, the data from which are typically not able to be aggregated,
3. The AAC&U VALUE rubrics can be aggregated but they are either too general for valid and reliable application directly at the course assignment level or too labor intensive to mark separate from course grading, and
4. The need to build instructor experience and capacity for assessment of course assignments in a way that is aligned to VALUE rubric criteria.

The purpose of BASICS is to generate assignment level rubrics that are practicable, discipline-specific articulations of the larger (VALUE) rubric. To enable this, BASICS is designed to:

1. Make the assessment of cognitive skills accessible for instructors who may be novices in criterion-based assessment, and help them provide task-specific language that clearly defines expectations for their students, and

2. Retain the hierarchy and taxonomy necessary for comparing student achievement between groups, or longitudinally (allowing for data aggregation).

METHOD

Part of the LOA project involved identifying where cognitive skills were being demonstrated in courses, such that student work samples could be evaluated. This began with discussions with course instructors to investigate alignment between existing assessment tasks and the dimensions of the VALUE rubrics. It was apparent that the guidelines for assignments did not often specifically require students to demonstrate cognitive skills. The LOA project researchers offered the VALUE rubrics to instructors as prompts for adapting assignments and assessment material in courses, but this was a difficult process for instructors to accomplish on their own because of the very general nature of the VALUE rubric criteria. Where appropriate, research personnel worked with instructors one-on-one to identify and develop methods for assessing learning outcomes and help refine particular approaches. However, this approach was labour intensive (took time and expertise), and was not generalizable beyond the individual instructor. Therefore, the challenge for the project was to develop sustainable methods for developing and assessing cognitive skills in a way that authentically reflected what instructors wanted their students to demonstrate in their courses. BASICS was designed to create assignment level rubrics for assessment of course deliverables such as analysis and research projects, design projects, investigations or structured inquiries.

DESCRIPTION OF THE TOOL

BASICS is a learning-focused interactive web application used by faculty or instructors to describe critical thinking, creative thinking or problem solving outcomes for students. BASICS is grounded in the AAC&U's VALUE rubrics, so the data generated via the disciplinary-specific rubrics produced by BASICS may be able to be aggregated across a program or an institution. The tool is comprised of a workflow that breaks down the language of high-level descriptors in the VALUE rubrics and automates an editable assignment level rubric. There were seven stages involved in the development and testing of the rubric builder (see Table 1).

Figure 1. Five-step process for Building Assessment Scaffolds for Intellectual Cognitive Skills
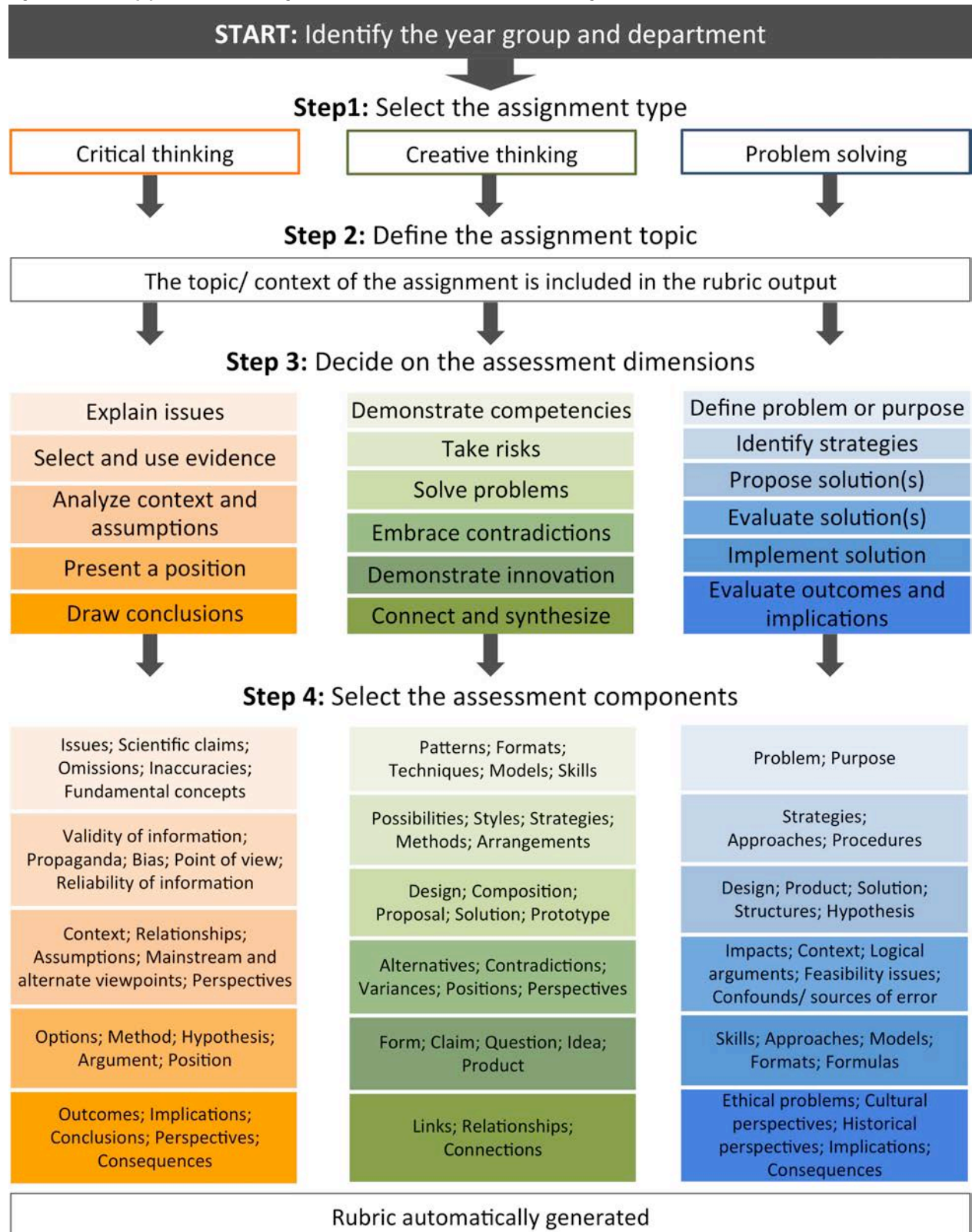
START: Identify the year group and department

Step 1: Select the assignment type

Critical thinking	Creative thinking	Problem solving

Step 2: Define the assignment topic

The topic/ context of the assignment is included in the rubric output

Step 3: Decide on the assessment dimensions

Critical thinking	Creative thinking	Problem solving
Explain issues	Demonstrate competencies	Define problem or purpose
Select and use evidence	Take risks	Identify strategies
Analyze context and assumptions	Solve problems	Propose solution(s)
Present a position	Embrace contradictions	Evaluate solution(s)
Draw conclusions	Demonstrate innovation	Implement solution
	Connect and synthesize	Evaluate outcomes and implications

Step 4: Select the assessment components

Critical thinking	Creative thinking	Problem solving
Issues; Scientific claims; Omissions; Inaccuracies; Fundamental concepts	Patterns; Formats; Techniques; Models; Skills	Problem; Purpose
Validity of information; Propaganda; Bias; Point of view; Reliability of information	Possibilities; Styles; Strategies; Methods; Arrangements	Strategies; Approaches; Procedures
Context; Relationships; Assumptions; Mainstream and alternate viewpoints; Perspectives	Design; Composition; Proposal; Solution; Prototype	Design; Product; Solution; Structures; Hypothesis
Options; Method; Hypothesis; Argument; Position	Alternatives; Contradictions; Variances; Positions; Perspectives	Impacts; Context; Logical arguments; Feasibility issues; Confounds/ sources of error
Outcomes; Implications; Conclusions; Perspectives; Consequences	Form; Claim; Question; Idea; Product	Skills; Approaches; Models; Formats; Formulas
	Links; Relationships; Connections	Ethical problems; Cultural perspectives; Historical perspectives; Implications; Consequences

Rubric automatically generated

Step 5: Edit rubric scaffold to semantic preferences and finalize

TABLE 1. Rubric-builder development stages.

DEVELOPMENT STAGE	*DESCRIPTION*
Rating samples	Course artefacts (assignments) were collected from consenting students from 12 different courses in Engineering disciplines, Drama, Psychology, and Physics, from 1st to 4th year (n= 949). Research assistants used the VALUE rubrics to mark the assignments and make annotations noting the evidence for the demonstration of each of the assessment criteria.
Synthesizing common elements- generating assessment components	The annotations were synthesized to generate a list of common elements within critical thinking, problem solving, and creative thinking at particular levels. The assessment components, and semantic arrangements of the rubric descriptors were based on these annotations.
Developing a workflow	The rubric development workflow was framed by the processes used in one-to-one meetings between instructors and research personnel within the LOA project. The processes were simplified for online use into five steps, aimed at identifying outcomes that align with a course assigned task.
Creation and testing	The information technology department (IT) designed a user interface for the five-step workflow, providing the common elements as choices, and using the semantic arrangements to automate the rubric generator. A database was created to enable the retrieval and search functions. The resulting rubric builder was named BASICS (Building Assessment Scaffolds for Intellectual Cognitive Skills), and functions in both desktop and mobile formats.
Piloting the rubric builder	Instructors from a range of disciplines across various institutions were invited to test the functionality of the tool and provide feedback. The feedback was used for refinement of language and functionality.
Evaluating validity	Evaluation of rubrics produced using BASICS was conducted based on a common set of criteria (see Dawson, 2015).
Investigating reliability	Course assignments were marked by Teaching Assistants, using a BASICS developed rubric, and then compared with independent VALUE rubric ratings made by research assistants.

Figure 1 provides an overview of the steps and options available in the BASICS rubric-builder. At the "start" prompt, instructors identify the discipline, and intended year group. The year group selected will determine the hierarchical level of the rubric (see Table 2). The levels are labeled as *developing*, *accomplished*, or *advanced*. A full user guide for instructors is available on the BASICS launch page: http://www.queensu.ca/qloa/assessment-tools/basics/.

Table 2 maps the VALUE rubric hierarchy to the BASICS rubric levels for each year group. In BASICS step 1, users choose critical thinking, creative thinking, or problem solving as the focus of their assignment. The assignment is described in step 2, the relevant dimensions, and rubric components are chosen (steps 3 and 4), then BASICS combines the selected options to produce an assessment rubric. Step 5 is intended for reflection on the suitability and specificity of the rubric for the instructor's purpose in assessing student learning. In other words, do the language and criteria describe the intended outcomes for the assignment? Using the editable fields in step five, instructors can fine-tune the rubric to clarify their expectations for students, thus facilitating constructive alignment of teaching, learning and

assessment (Biggs & Tang, 2011). Depending on the combination of elements chosen, there are over 6000 possible rubric variations that may be derived using the BASICS rubric builder. An example of one possible combination is shown in the rubric below (Table 3) and all of the BASICS rubrics generated thus far are archived here: http://www.queensu.ca/qloa/assessment-tools/basics/list.php. There is a search function on the page that enables rubrics to be located by topic, institution, department or year group.

Table 2. VALUE rubric levels mapped to the BASICS rubrics per year group

Year group selected on the BASICS rubric builder	VALUE rubric level				
	Not Demonstrated	Benchmark 1	Milestone 2	Milestone 3	Capstone 4
First year (Freshman)	Developing	Accomplished	Advanced		
Second year (Sophomore)		Developing	Accomplished	Advanced	
Third and fourth year (Junior and Senior)			Developing	Accomplished	Advanced

Table 3. Course rubric developed using the BASICS rubric builder

CRITICAL THINKING/ PROBLEM SOLVING	DEVELOPING (1 POINT)	ACCOMPLISHED (2 POINTS)	ADVANCED (3 POINTS)
Issues	Presents issues and omissions in study design and statistical analysis used.	Critiques key issues and omissions impacting the study design and statistical analysis used.	Debates key issues and omissions impacting study design and statistical analyses used.
Context/ Assumptions	Explains contexts and assumptions as they relate to the study design and statistical analysis used.	Analyses contexts and assumptions and their impact on the study design and statistical analysis used.	Synthesizes contexts and assumptions to objectively explore their impact on the study design and statistical analysis used.
Implementation	Uses simple routine procedures when applying approaches to solve the issues presented.	Adapts procedures when applying approaches to solve the issues presented.	Tailors procedures when applying approaches to solve the issues presented.
Implications/ Conclusions	Explains implications	Addresses implications, making recommendations for corrections or future work	Weighs impacts, accounting for implications when recommending corrections or future work.

EVALUATING THE RUBRIC

When evaluating an assessment tool, we need to consider whether the rubric measures the intended learning outcome (validity) and whether that measurement can be consistently applied to the

assessment context (reliability). Dawson (2015), proposes a rubric design element evaluation framework comprised of 14 categories with criteria recommended for consideration when designing assessment rubrics. Table 4 describes each of these criteria specific to the BASICS rubric's content. See Dawson (2015), for a detailed explanation of each of the categories and the literature surrounding each.

TABLE 4. BASICS rubric design elements evaluation

ELEMENT	BASICS RUBRIC EVALUATION
Specificity	The scaffolds are described as "task specific" because they are designed to cater to a particular assignment. A more suitable description might be topic specific, as the evaluation criteria describe assessment of skills within a context rather than being limited to the specific task.
Secrecy	Not prescribed. It is up to the individual instructor whether they share the rubric prior to the assessment or provide it with feedback to the student when the work is marked.
Exemplars	No work samples are provided with the BASICS rubric tool.
Scoring strategy	There is no specific scoring strategy provided, but the performance levels are designed for use as an analytical rubric to arrive at marks and grades.
Evaluative criteria	No specific attributes are provided in the template, but the edit function offers the opportunity for these to be added if required.
Quality levels	There are five possible levels presented. These align hierarchically with the four VALUE rubric levels of one to four, with the addition of a level zero.
Quality definitions	The expected performance quality levels are given with the labels developing, achieved, and advanced.
Judgment complexity	It is an unstated expectation that evaluative expertise and marker training should be provided to users of the rubric.
Users and uses	The rubric tool is intended for use by instructors to develop course assessment materials. The rubrics themselves may be used by students to determine expectations, and would be used by instructors or teaching assistants to evaluate student work, and to provide feedback to students. In addition, the application also provides a database of previously generated rubrics that are searchable by topic, department, or year group.
Creators	Course instructors are the intended creators of a rubric, but it would be feasible for students to be active in the rubric construction process.
Quality processes	No formal approaches to ensure the reliability and validity of the rubric are provided.
Accompanying feedback information	The rubrics are intended for in-class use to make performance expectations clear and provide feedback on student performance.
Presentation:	Rubrics display with colored bands for each of the dimensions shown in the left hand column, and the three levels of descriptors elaborated from left to right. The top row provides a header indicating levels of *developing*, *accomplished* and *advanced*.
Explanation	Instructions or other additional information provided to users (see Simper, 2016).

USING BASICS FOR ASSESSMENT

A pilot study was undertaken to investigate the relationship between assessments conducted for course grading and the external assessment of the same artifacts using the VALUE rubrics. Psychology students in Advanced Statistical Inference (PSYC 301) completed a diagnostic assignment for course credit. Recruitment was conducted as per ethical approval, by a research associate, during the weekly lab sessions. There were 126 students in the class, 93 of these attended their specified lab session during the recruitment week. Of these students, 63 consented to have their assignments used for research purposes. The diagnostic assignment instructed students to critically assess a research article in order to identify the statistical and methodological issues presented within it. Students were instructed to submit independent and originally constructed assignments in the form of a short paper (500 words or less), worth 3% of their final grade. The article was critically reviewed with the awareness that the paper had been pulled from publication due to inaccuracies and flaws in statistical methods.

The course instructor generated a four-dimension rubric using the BASICS tool (see Table 3). Two of the dimensions related to critical thinking (*explanation of issue*, and *influence of context and assumptions*) and two of the dimensions related to problem solving (*implement solutions* and *evaluate outcomes*). The rubric was applied analytically, with a three-point scale for each of the four dimensions. A maximum of 12 points was attainable for an assignment that earned all *advanced* levels. The three levels on the course rubric aligned with the upper three levels of the VALUE rubric (see mapping for third year in Table 2). When compared directly with the VALUE rubric dimensions, the course-based rubric descriptors provided more specificity, but less breadth than the corresponding VALUE rubric descriptors.

The course was comprised of five lab sections with each section assigned a teaching assistant (TA). To observe the 'in the field' application of the BASICS generated rubric, researchers did not intervene in any way in the course marking process. The course instructor trained the TAs in the use of the rubric, and the five TAs individually marked the assignments for their specific lab sections. Two research assistants (RAs) were trained in the use of the VALUE rubrics (see steps below), and both RA's rated each (n=63) of the work samples (assignments). The process for using the VALUE rubrics involved three steps: (1) building a common understanding, (2) rating work sample, and (3) calibrating ratings. The "markers" referred to here are the research RA's. The three steps of the VALUE Rubric process framework occurred as follows:

1. Building a common understanding
   a. Read through the assignment brief and example responses to build an understanding of the nature and context of the course assignment,
   b. What were the students directed to demonstrate? (This required reading the research paper that the students were responding to), and
   c. Operationalizing the "issues," "contextual factors," and "assumptions" relevant to the student responses.
2. Rating work sample
   a. Collectively work through a single student response (not one included in the research sample) to identify evidence for each of the dimensions to be rated. RAs would then discuss what level the evidence suggests the response is demonstrating,
   b. Individually rate 5-10 work samples at a time, compiling an annotated list to back up the decision for each of the criteria, and
   c. Assign a performance level (for each dimension) for the work samples.

3. Calibration
   a. The two markers use their annotations to discuss any differences between levels assigned.
   b. In some cases this process results in one or the other of the markers adjusting their level on a dimension. The rating process is based on individual interpretation, so differences in level determinations were occasionally observed.
   c. Repeat steps 2b- 3 for the remainder of the work samples. Generally, the greater the number of assignments that are rated, the fewer differences there are in ratings.

RESULTS

The inter-rater reliability was investigated; the pre-calibration and post-calibration Kappa scores are displayed in Table 5. Prior to calibration there was 70% exact agreement, after calibration the agreement was 95%. The TA scores on the (BASICS) course rubric were numerically recoded to correspond with the VALUE rubric levels for comparison of means, i.e. a score of 1 on the course rubric became a score of 2 to correspond with the intended level on the VALUE rubric; 2, became 3; and 3 became 4. Descriptive statistics for the five course TAs and the two research assistants, as well as a measure of internal consistency are presented in Table 6. The internal consistency of the TAs and RA's was (α=.64 to .90) with the exception of TA_2 (α= -.11).

TABLE 5. Cohen's Kappa agreement between RA1 and RA2

	n	EXPLANATION OF ISSUES	CONTEXT AND ASSUMPTIONS	IMPLEMENT SOLUTIONS	EVALUATE OUTCOMES
RA Pre-calibration (κ)	63	.81**	.74**	.60**	.65**
RA Post-calibration (κ)	63	1.00**	.98**	.89**	.93**

**p<0.01

Table 6. Descriptive statistics for RA raters and teaching assistants.

		EXPLANATION OF ISSUES		CONTEXT AND ASSUMPTIONS		IMPLEMENT SOLUTIONS		EVALUATE OUTCOMES		CRONBACH'S ALPHA
	n	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	α
RA_1	63	2.30	.66	2.30	.89	2.52	.71	2.67	.94	.84
RA_2	63	2.30	.66	2.31	.89	2.48	.71	2.69	.94	.83
TA_1	12	3.17	.72	3.42	.70	3.58	.67	3.25	.622	.84
TA_2	9	3.80	.42	2.70	.68	3.10	.88	2.60	.84	-.11
TA_3	17	2.71	.59	2.59	.62	2.82	.73	2.71	.69	.89

TA_4	8	3.38	.74	3.38	.52	3.25	.71	3.38	.74	.64
TA_5	16	3.19	.75	3.06	.68	2.94	.77	3.25	.68	.90
$RA_{average}$	63	2.35	.74	2.33	.91	2.51	.74	2.68	.91	.86
$TA_{average}$	63	3.17	.73	2.98	.71	3.10	.78	3.02	.75	.82

The scores for TA_2 were further investigated and negative inter-item correlations were found between dimensions, this result was anomalous with other TA and RA scoring. Due to the inconsistency of TA_2's scoring, the nine scores from this lab section were excluded from further analyses. The scores assigned by TAs were generally higher than the RAs for each of the dimensions, but had less variance. To investigate possible reasons for this, consideration was given to the specificity of the descriptors, the level of expertise, the objectivity of the scorers, and the calibration of scores (see discussion below).

The relationship between the BASICS and VALUE ratings was investigated to provide an estimate of concurrent validity. Supported by high agreement between the RA's ratings, a VALUE dimension average was calculated to reduce the RA ratings to a single score for each of the samples on each dimension. The course assignments included in the analyses were scored with a degree of consistency by four separate TAs. The TA scores were entered in a series for correlation analyses to enable the direct comparison between the assignment scores for VALUE and BASICS rubric dimensions. Pearson's correlations of between .45 and .61 (see Table 7) suggested that there was a significant similarity between the criteria in the BASICS and VALUE rubrics, providing support for concurrent validity.

TABLE 7. Correlations between averaged RA scores and TA scores (n=54) for each category assessed.

RUBRIC DIMENSION	PEARSON'S CORRELATION	INTRACLASS CORRELATION	CONFIDENCE INTERVAL
	r	ICC	95% CI
Explanation of Issues	.45[**]	.61[**]	[.33, .78]
Context and Assumptions	.49[**]	.65[**]	[.40, .80]
Implement Solutions	.49[**]	.68[**]	[.45, .82]
Evaluate Outcomes	.61[**]	.68[**]	[.45, .82]

[**] significant at the 0.01 level.

Two-way mixed Intraclass correlations (ICC) were calculated between the RAs average and TAs scores to investigate the reliability (consistency) between the BASICS rubric and VALUE ratings. The ICC indicated moderate consistency under all dimensions, suggesting that even though the marking was undertaken by different raters the resulting scores from the BASICS rubric were moderately consistent with those on the VALUE rubric dimensions. Entered collectively into the analysis, average measures yielded a high degree of consistency between VALUE and BASICS measures. ICC was .90, with a 95% confidence interval from .85 to .93 $[F(53,371)= 9.60, p< .001]$.

INSTRUCTOR FEEDBACK

Ethics approval was granted to pilot the tool nationally and internationally and instructors were recruited at a meeting of undergraduate faculty leaders and at academic conferences in Ontario, Canada, and Melbourne, Australia. The pilot group was composed of 45 instructors from 21 institutions across Australia and Canada. Nine were from the College sector and 21 were from Universities. The departments included education, sciences, social sciences, humanities, and the arts. During the pilot, ten of the instructors provided written feedback (eight from Canadian institutions, two from Australian institutions). The feedback survey had six questions; two text responses relating to the language and functionality of the tool, and three Likert scale questions asking about the ease, usefulness, and the likelihood of using the tool to support course-based assessment. There was a final open response question for general comments. Eight of the ten respondents (80%) reported that the tool was easy, or very easy to use; 70% reported the tool to be useful or very useful; and 50% reported that it would be likely, or very likely they would use the tool for course-based assessment. The comments were generally very positive, such as, "this was great"; "the design of it made it easy and quick to use, which is essential for faculty"; [the tool] "supports approaches to learning outcomes assessment…. [I can] work from the implicit goals to make them explicit, and then work to make them measurable"; and "BASICS was greatly beneficial to me (because) it gave me descriptors that allowed me to explain to my teaching assistants what was expected, and to provide feedback to the students." Not all instructors found the tool useful, with one respondent saying that (BASICS) "definitely got me thinking, but I don't think I'd use it myself."

DISCUSSION

The goal of BASICS was to make assessment of cognitive skills easier for instructors, and to enable aggregation of data. Using the VALUE rubrics in their original form for course assessment was not an option because without adaptation or training they were too general for direct application. Instructor feedback from the pilot suggested that BASICS was helpful for instructors to identify learning outcomes, and made it easier for them to assess elements of student learning. For assessment data derived from rubrics constructed with BASICS to be aggregated, as data from the VALUE rubrics is, consistency between the two forms is necessary.

The pilot provided comparison between ratings of five TA's and two RA's. The two measures had similar reliability, but the RA's required considerable training to apply the very general descriptors to the course assignment. The BASICS rubric was applied with limited training, yet there was moderate consistency between individual BASICS and VALUE rubric dimensions (see Table 7), with a high degree of consistency for the combined average measures. Three main issues arose from using data derived from TA marking conducted outside a controlled setting; namely (a) TAs marked using the BASICS criteria and RAs used the VALUE criteria, (b) researchers not having control of the TA marking procedures, and (c) that the TAs knew whose assignment it was they were marking. These issues are expanded on here:

1. Error would be minimized if it were possible to have either the TA's or RAs conduct both the BASICS and VALUE ratings. This was not possible because having already rated using one measure, ratings using the second measure would be inherently biased by the first.
2. The TA's and RA's have different levels of training, and the RA's procedures included the process of calibration. Using data derived through regular course marking was seen as a measure of authenticity but the union constraints mean that there are a maximum number

of hours that a TA can work. To investigate the use of the rubric 'in the field', researchers did not manipulate or control how the course assessment was conducted. Inconsistencies in TA ratings (specifically TA$_2$) could have resulted from inexperience in assessment. TA training is not the topic of this paper, but if TA's were required to identify specific evidence within the work sample to justify the level they awarded, it may be that inconsistencies could be avoided. If this procedure is not observed, then the ratings can become subjective. In addition, periodically seeking consensus about judgments ensures consistency between ratings. This process, called calibration, was undertaken by RA's but not by TA's. Although the TA's were trained, over time they may have been making individual interpretations about assessment criteria and their relation to the course assignment. The RAs had considerable practice in using the VALUE rubrics and calibrated their scores periodically.

3. The scores assigned by TAs were generally higher than the RAs. A potential source of error could have been a halo effect (Fox, Bizman, & Herrmann, 1983) from course TAs. When TAs were responsible for a specific group of students, they may have built rapport, becoming familiar with individuals in their class. This familiarity may have led the TA's to view the student submissions with an overly positive perspective. The RA's were able to rate work objectively; they did not know the identity of the students who contributed work samples and were impartial to the resulting ratings.

In addition to the above issues, there were differences in specificity of descriptors in BASICS compared with the VALUE rubrics. There were moderate correlations (not high); the rubric criteria were similar but differed in specificity. The VALUE rubric descriptors are very general and are designed to be applied to a range of assignments across the institution, whereas the wording of the BASICS rubric criteria was specific to a single assignment. For example, the course rubric descriptor for the highest level of *explanation of issues* was "debates key issues and omissions impacting on study design and statistical analyses used," whereas the corresponding VALUE rubric descriptor states "issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding." This particular VALUE criterion would have been difficult for students to demonstrate because there was a maximum word count limiting the comprehensiveness of the response. The assignment used in this study prompted students to demonstrate components of critical thinking and problem solving. Regular and ongoing assessment of these skills would be necessary to provide additional opportunities for students to demonstrate the full range and complexity of the assessment dimensions.

It is now well established that "students need a broad set of essential skills and abilities in addition to a strong knowledge base to achieve success in today's global society" (T. Rhodes, 2010, p. 14). For institutional demonstration of these skills, course assessment needs to be aligned with common criteria. The secondary intention for BASICS was to build structures to retain the hierarchy and taxonomy necessary for comparing student achievement between groups, or longitudinally, enabling aggregation of data. On face value, the hierarchy is present, but further research would need to be undertaken to ascertain whether reliabilities demonstrated here are replicable for different raters, for alternate assessment tasks, and across a range of disciplines.

LIMITATIONS

BASICS is limited to developing rubrics for cognitive skills; inferences should not be made that discipline specific knowledge is not important. For evaluation of specific student knowledge, additional

dimensions would need to be included in the rubric. The main limitation of the pilot was that it utilized data from five TA's and two RA's based on a single course assignment. There is a need for further investigation of the use of BASICS rubrics in multiple contexts and across different learning areas. The initial data mentioned in this paper suggested that the BASICS rubric has merit regarding validity and reliably, but more data are needed to validate preliminary conclusions presented here.

BASICS was not designed to be a complete solution to assessing cognitive skills. Also, simply creating a rubric should not be expected to have a direct impact on student learning. As user feedback suggests, BASICS can be helpful for identifying desired assessment criteria. In this way, it becomes part of a backward design process (Wiggins & McTighe, 2005). Assessment rubrics are a key piece of a package that includes, among other things, prompting student learning through engaging assessment tasks and providing effective learning pathways. Timmerman et al., (2011) identified alignment as an issue with using the VALUE rubrics. By using BASICS, the instructors in the pilot were able to develop their own rubric for a specific course assignments. In the generation of the rubric, the instructor had the opportunity to review the task guidelines to ensure that the skills were elicited, thus avoiding the alignment issue common to the use of VALUE rubrics. The goal in the future would be to include a notes section along with an upload function, where instructors could upload the course assignment for which the rubric was designed. This would enable an "assignment library" similar to the one compiled by The National Institute for Learning Outcomes Assessment (NILOA) for the Degree Qualifications Profile (DQP) (see "Assignment Library," n.d.).

CONCLUSIONS

Evidence provided here suggests that the BASICS rubric authoring tool (Home | Queen's Learning Outcomes Assessment, n.d) is a promising device for supporting instructors in developing rubrics for evaluation of student performance of cognitive skills, with the goal of enabling aggregation of data across the institution. Instructor feedback was positive, and although limited in scope, analysis of BASICS rubric scores found high internal consistency. Investigation of intraclass correlation coefficients between the BASICS rubric developed in the pilot and the corresponding VALUE rubric dimensions indicated moderate consistency, suggesting that BASICS may be an alternative to direct use of VALUE rubrics for assessment across an institution. There were some systematic differences between the BASICS and VALUE rubric scoring; the BASICS scores were consistently higher than the VALUE rubric scores. The differences were thought to be due to either the specificity of the language, the level of expertise of the marker, or the level of impartiality in scoring. The structure of the BASICS rubric authoring tool was grounded in the AAC&U VALUE rubrics; therefore, data aggregation is feasible, but more work needs to be done to investigate the reliability of data aggregated from BASICS rubric assessment. The research team is continuing to investigate the application of BASICS rubrics in a range of alternate contexts, and encourage others to trial the use of BASICS for assessment of cognitive skills at their institution.

Natalie Simper is the Assessment Research Projects Manager at Queen's University in Canada. She comes from an Australian Senior-Secondary/ Post-Secondary teaching background, with experience at the Statewide level in curriculum development, large-scale assessment, and evaluation and assessment of outcomes based education.

REFERENCES

Assignment Library. (n.d.). Retrieved July 14, 2016, from http://www.assignmentlibrary.org/search

Association of American Colleges and Universities. (2014, October 7). VALUE [Website]. Retrieved September 30, 2016, from https://www.aacu.org/value

Bernstein, D., & Greenhoot, A. F. (2014). Team-designed improvement of writing and critical thinking in large undergraduate courses. *Teaching & Learning Inquiry: The ISSOTL Journal*, *2*(1), 39–61.

Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University. Maidenhead: Open University.* UK: Mc-Graw-Hill Education.

Boud, D. (1995). Assessment and learning: contradictory or complementary. In *Assessment for learning in higher education* (pp. 35–48). Kogan Page.

Brown, S. (2004). Assessment for learning. *Learning and Teaching in Higher Education*, *1*(1), 81–89.

Clark, J. E., & Eynon, B. (2011). Measuring student progress with e-portfolios. *Peer Review*, *13/14*(4/1), 6.

Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, (ahead of print), 1–14.

Frank, B., Simper, N., & Kaupp, J. (2016). How We Know They're Learning: Comparing Approaches to Longitudinal Assessment of Transferable Learning Outcomes (pp. 1–14). Presented at the 2016 ASEE Conference and Exposition, New Orleans, LA: ASEE Conferences. https://doi.org/10.18260/p.25493

Fox, S., Bizman, A., & Herrmann, E. (1983). The halo effect: Is it a unitary concept? *Journal of Occupational Psychology*, *56*(4), 289–296.

Gray, P. J. (2013). Developing Assessment Rubrics in Project Based Courses: Four Case Studies. In *9th International CDIO Conference, Cambridge, Massachusetts*. Retrieved from http://ww.cdio.org/files/document/file/W5A4_Gray_006.pdf

Herbert, S. (2015). *Focus on Outcomes, Centre on Students: Perspectives on Evolving Ontario's University Funding Model*. Ontario, Canada: Ministry of Training, Colleges and Universities. Retrieved from http://www.tcu.gov.on.ca/pepg/audiences/universities/uff/UniversityFundingFormulaConsultationReport_2015.pdf

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130–144.

Kuh, G. D., Jankowski, N. A., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in US Colleges and Universities.* Urbana, IL: University of Illinois and Indiana University: National Institute for Learning Outcomes Assessment (NILOA). Retrieved from http://www.learningoutcomeassessment.org/documents/2013%20Survey%20Report%20Final.pdf

Partnership for 21st Century Skills. (2008). [PDF document]. Retrieved July 12, 2016, from http://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf

Queen's University. (2014, October 7). Home | Queen's Learning Outcomes Assessment [Website]. Retrieved December 22, 2015, from http://www.queensu.ca/qloa/

Rhodes, T. (2010). Since we seem to agree, why are the outcomes so difficult to achieve? *New Directions for Teaching and Learning*, *2010*(121), 13–21.

Rhodes, T. L. (2011). Emerging evidence on using rubrics. *Peer Review*, *13*(4/1), 4–5.

Rhodes, T. L., & Finley, A. P. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment*. Washington, DC: Association of American Colleges and Universities.

Siefert, L. (2011). Assessing general education learning outcomes. *Peer Review*, *13*(4/1), 9–11.

Simper, N. (2016, July). Authentic Assessment of Cognitive Skills. Retrieved February 10, 2018, from http://www.queensu.ca/qloa/authentic-assessment-cognitive-skills

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for

assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509–547.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by Design*. Alexandria, VA: Association for Supervision and Curriculum Development.