2019

Validating Bloom's Revised Taxonomy as a Rubric for Assessing Middle School Students' Levels of Thinking

Siri Torrence DeForest Reynolds
Walden University

Walden University

College of Education

This is to certify that the doctoral study by

Siri DeForest Reynolds

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Jennifer Seymour, Committee Chairperson, Education Faculty
Dr. Christopher Godat, Committee Member, Education Faculty
Dr. Richard Hammett, University Reviewer, Education Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Validating Bloom's Revised Taxonomy as a Rubric for Assessing Middle School Students' Levels of Thinking

by

Siri DeForest Reynolds

MA, University of Phoenix, 2008

BS, The College of Saint Rose, 2003

Project Study Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

June 2019

Abstract

Educators in a rural charter middle school in the United States were challenged with the reliable assessment of student thinking skills even though the development of higher order thinking was an espoused goal for the school. The purpose of this study was to validate a new rubric based on Bloom's Revised Taxonomy (BRT) to reliably assess student levels of thinking as reflected in the students' written work. A quantitative, nonexperimental design was used. The focus of the research questions was on the BRT rubric's reliability and validity. Interrater reliability was assessed using Krippendorff's alpha. Validity was explored by assessing the relationship between the BRT scores collected in this study to the original teacher scores of students' archived writing samples. Reliable, unrelated scores would have suggested that the two processes were scoring different constructs. The convenience sample of 8 volunteer teachers scored papers using the new BRT rubric. Each teacher scored 52 writing samples, 2 each from 26 students in the 7th grade. The Spearman correlation coefficient between the BRT and original teachers' scores was not statistically significant. The teachers' original scores could not validate the BRT as a measuring tool. Also BRT measure failed to demonstrate evidence of reliability (Krippendorf's $\alpha = .05$). A position paper was created to present the results of this study and to explore possibilities for improving the assessment of thinking. Positive social change may be encouraged by the use of a reliable and valid scoring process to quantify levels of thinking. A reliable scoring process for levels of thinking could lead to more balanced curricula, instruction, and assessment ultimately providing a base for customized student learning experiences.

Validating Bloom's Revised Taxonomy as a Rubric for Assessing Middle School Students' Levels of Thinking

by

Siri DeForest Reynolds

MA, University of Phoenix, 2008

BS, The College of Saint Rose, 2003

Project Study Submitted in Fulfillment

of the Requirements for the Degree of

Doctor of Education

Walden University

June 2019

Dedication

This study is dedicated to Dorothy Bachelor DeForest. She is my grandmother, was my best friend, and a million things to a million people. Rest in peace.

Acknowledgments

First off, I would like to thank Dr. Jennifer Seymour. Without her guidance and steadfast support this work would be unfinished. Thank you for believing in my “big ideas,” and thank you even more for helping me shape them into something worthwhile! I must also thank my family for tolerating my perpetual schooling.

Table of Contents

List of Tables

Section 1: The Problem

The Local Problem

According to educational experts, students in the United States are ill-prepared to face uncertainty and develop solutions for 21st century challenges, many of which are still unidentified. In order to be prepared, students must learn higher-order thinking skills (HOTS; Heick, 2016). There are some assessments for HOTS, but they are unwieldy, and the scoring is subjective (Silvia et al., 2008). The problem for this study is that for HOTS there is no assessment tool that can be used regularly, efficiently, and reliably (Brookhart & Chen, 2015). In response to this need, I investigated the discriminant validity and interrater reliability of Bloom's Revised Taxonomy (BRT; Wilson, 2013), a quantitative categorical scoring taxonomy, as an assessment rubric for HOTS.

BRT is currently a widely accepted taxonomy for evaluating the existence of a continuum of lower order thinking skills (LOTS; Yassin, Tek, Alimon, Baharom, & Ying, 2010). Although some researchers (Hess, Jones, Carlock, & Walkup, 2009; Thompson, Luxton-Reilly, Whalley, Hu, & Robbins, 2008; Yassin, Tek, Alimon, Baharom, & Ying, 2010) have used BRT for measuring LOTS as a rubric, they have not evaluated the validity and reliability of using BRT as a rubric to guide assessment, according to my review of the literature. My purpose, therefore, was to evaluate the validity and reliability of using BRT as a rubric based on my review of literature in the field supporting the need for continued evaluation of the assessment of thinking capacity as evidenced in recent literature (Lo, Larsen, & Yee, 2016; Pecka, Schmid, & Pozehl, 2014). Anticipated implications for positive social change include improving teachers'

abilities to assess and teach HOTS, thereby helping to close the achievement gap between the U.S. education system and its global competitors (Wiliam, 2011).

Assessment is a global problem with local implications. Currently, the United States lags in comparison with other nations in achievement on international assessment charts (Comparative & International Education Society, 2014). More concerning than the low scores on standardized tests is that current standardized tests measure rote knowledge and content awareness and do not address assessment of HOTS and other 21st century capabilities, according to Koehler, Mishra, Kereluik, Shin, and Graham (2014). This project study is a call to action for the development of an alternative assessment to the widely used standardized assessment tools that currently prevail (Benjamin et al., 2012). Using BRT as an assessment taxonomy could extend educational focus from increasing student content knowledge to developing student thinking skills along a continuum from LOTS to HOTS, which would improve the education of HOTS.

At the local level, middle-grade teachers at the western U.S. charter school struggle with the lack of a rubric to assess students' ability to think at a range of levels. The problem is that teachers do not have a reliable tool for assessing student writing for evidence of HOTs. The school is dedicated to teaching HOTS but lacks an assessment tool to evaluate the acquisition of these skills. In addition, educators at the school use assessment software that collects a sizable amount of written student products, but they lack a means to evaluate student thinking demonstrated in each assignment.

Using the BRT as a rubric to evaluate HOTS could be a viable solution to fill this local gap in practice of inefficient and unreliable assessment of students thinking as

produced in writing. The BRT rubric may thus provide a consistent school-wide criterion against which to evaluate student thinking. Ultimately, the BRT rubric could be useful to track cognitive growth across time. It is important to note that this project study is constrained to evaluating the validity and interrater reliability of BRT as a thinking assessment rubric at the middle school level.

The gap in practice at the local level was the lack of adequate assessment of student thinking. The charter school promotes instructional practices that foster HOTS without any evidence of success. The existing formative and summative assessments only aim to capture low-level knowledge recall, even though HOTS are a focus of the charter school. The intention to teach HOTS is evident, but work remains to integrate the assessment of the HOTS.

The primary assessment of low-level thinking was built into the instructional plan, even though the curriculum largely supports assisted learning environments instead of content lecturing. Additionally, according to the administrators at the school the charter's curricula generally lack accountability through formative assessment data. The charter school could have focused on gathering data throughout learning experiences to impact ongoing instructional design and implementation. However, the assessment plans only focus on the lower levels of student thinking; they neither account for the assessment of higher-order thinking nor the assessment of any range of student thinking. Different forms of assessment, such as the BRT rubric, may hold promise for improving teacher behaviors based on more accurate feedback about where students are in their development of HOTS. In summary, the learning environment was partially consistent

with best practices for reforming classrooms, but the assessments only documented LOTS, even though the intention was to teach HOTS.

Some of the impetus for this study derived from my experiences at a similar charter school. In the past, as part of my job responsibilities I once observed an educator in this charter school teach a lesson by focusing on higher-order thinking and prompting higher-level student discussions and analysis. For example, the teacher prompted students using a questioning technique in which she would provide students with a statement and then ask them to generate as many questions as possible. In responding to these questions, students would provide complex open-ended responses versus *yes* or *no* answers (see Rothstein & Santana, 2011).

The assessment planned by that teacher to document learning during the lesson was an interactive notebook, which is a collection of notes with content guided by a facilitator wherein students are supposed to create evidence of scaffolded learning by recording the exploration of thought, creation of connections, and active learning (Carter, Hernandez, & Richison, 2009). Based on the teacher and student discussion at the end of the lesson it was evident that the completed interactive notebooks included largely low-level content answers instead of the HOTS displayed in the classroom dialogue. Additionally, this notebook could only be scored for content because there was no tool for scoring students for different levels of thinking. The BRT rubric investigated in this study may be an effective tool for tracking the use and development of the different levels of HOTS.

Traditional content assessments are practical and logical, while the evaluation of HOTS using written assignments and open-response assessment items represents a new development of assessment (Holt, Young, Keetch, Larsen, & Mollner, 2015). The need for balancing content and thinking aspects of assessment instruments has been a topic of conversation at this charter school. Additionally, in the past, as part of my job responsibilities, I observed a dialogue amongst staff during the end-of-the-year review of normed testing data. For example, the staff repeatedly complained that the results from the interim data analysis of end-of-year review of normed testing data were superficial and did not connect to any forms of learning beyond general recall of content or summarization skills. Moreover, these colleagues observed that the components of the testing instruments provided only snapshots of a student's ability to read and select from multiple-choice answers.

My colleagues were dismayed that there was no assessment based on performance to measure levels of thinking. In all, the staff voiced the desire for an assessment of thinking that they could use to view students' varied levels of thinking. The lack of this type of assessment stems largely from the subjective nature and sizable amount of time involved in using current methods to reliably measure student thinking (Yan & Cheng, 2015). It would be beneficial, according to my colleagues, to use a more objective measure that could be used on a larger data set, such as the BRT could potentially be used.

The first instrument employed by the district was the Standardized Test for the Assessment of Reading , an assessment that yields normed, archival data (Renaissance,

2019). The second instrument is the Colorado Measures of Academic Success, which yields descriptive information about school performance in reading, writing, math, and science (Colorado Department of Education, 2018). Colorado Measures of Academic Success yields broad data related to reading, writing, math, and science (Colorado Department of Education, 2018). Meanwhile the Standardized Test for the Assessment of Reading yields student-level, with subdata broken down into specific deficiencies and strengths (Renaissance, 2019). A third standardized test that the school prepares its students for is the Partnership for Assessment of Readiness for Career and College (PARCC; Pearson, 2010) assessments in the areas of English language arts and mathematics. PARCC is a national standardized assessment correlated to the Common Core State Standards but is still in its infancy and therefore uncorrelated to existing standardized measures (Pearson, 2010). Neither PARCC nor the Standardized Test captures the demonstration of students' thinking as proposed in this study. Identification and tracking of the development of students' ability to think is something that is important to the teachers and administrators in this district, according to the school's headmaster, and a BRT rubric to assess LOTS and HOTS may contribute to solving this problem.

The absence of assessments that measure student thinking is the problem. On a large scale, assessment design does not indicate 21st-century learning goals that include thinking skills (Lamb, Marie, & Doecke, 2017). Students must demonstrate competencies in critical thinking, problem solving, collaboration, and autonomous independent transfer of knowledge to exercise higher order thinking (Wagner, 2014). Assessment data focused

on thinking and learning must regularly be collected, must inform instruction, and must be pulled from a pool of success criteria universal to the learning community (Brookhart & Chen, 2015; Moss & Brookhart, 2009). These success criteria may possibly be encompassed in the BRT rubric. Many scholars have discussed the need for better assessments; a recurring theme of research is the amount of time required to score and the subjectivity involved with scoring (Beck, 2006; Goldring et al., 2015). A BRT-based rubric may be quicker and more objective than the rubrics previously developed for a variety of assessments that do not assess levels of student thinking.

In response to the deficits businesses and colleges have identified, educators have begun the assessment of thinking on state standardized assessments. Such standardized assessment tools include PARCC and Smarter Balanced Assessments (SBA) in the form of Performance-Based Assessment components (Benjamin et al., 2012; Herman, Linn, & Moss, 2013). Considering that large-scale testing corporations have begun to focus on this area of need, the use of BRT as a rubric for the reliable assessment of thinking at the k-12 level is consistent with the direction of the field.

The subjectivity inherent in assessment of written work manifests through practice in many ways. Hess et al. (2009) noted the discrepancies in teacher scoring when teachers fall into old habits of scoring on academic enablers such as student past behaviors or achievements related to student work habits. For example, if a teacher scores an essay and the rubric is vague the teacher is likely to factor in historical subjective observations and associations from interactions with the student in the past (Brookhart & Chen, 2015). Additionally, a student's actual academic competence and habits may factor

into what a teacher identifies or focuses on if, for example, the student is frequently late in turning in work (Hess et. al., 2009). When scoring student writing the attitude of the teacher may be less open to possibilities of the higher range of LOTS to HOTS in each student's writing based on preexisting bias from interactions with students with poor academic habits.

Although a good deal of research is available related to classroom assessment, there is a gap in research around the documentation of student LOTS and HOTS (Wiggins & McTighe, 1998). McMillan (2013) identified the need to develop principles of assessment that document student learning, addressing specifically the necessity for a supporting body of research on classroom assessment. Furthermore, Brookhart (2013) discussed the need for developing evidence of in-depth descriptions of how teachers summarize and document learning and how learning progresses. Focusing on the 21st century, with the transformation in the contexts for assessment, Aagaard and Lund (2013) identified the lack of experience by educators in how to assess collaborative and interactively constructed learning (p. 223). There are projects to track groups of students' learning in addition to individuals. For example, Confrey and Maloney (2015) discussed the design of software to trace individual as well as collective learning trajectories. The collective learning environments must first be designed so that there are HOTS to track. One goal for this study, therefore, was to lay the foundation for tracking both group and individual HOTs by validating the BRT rubric for level of thinking.

There is a difference between design of learning questions and the assessment of the thinking generated from those questions. Although there are structures in place, such

as BRT, to guide the design of learning opportunities and questions that address higher order thinking, there is a breakdown in the assessment of the responses to the questions focused on higher order thinking (Vista, Care, & Griffin, 2015). In the design of evaluation tools to use in a formative or summative fashion, indicators must be identified for ideal student outcomes for specific tasks within a given discipline (Vandal, 2012). For example, Atherton (2013) discussed the phases of learning using a Structure of Observed Learning Outcomes taxonomy, in which indicators are checked off as the students' learning progresses through Piagetian developmental phases beginning with the prestructural through the extended abstract level in which students transfer from simple to complex applications. Following a developmental trajectory of learning from the LOTS to the HOTS is one long-term aim of the BRT rubric developed for use in this study. I sought to do the preliminary work of determining if the BRT can be used as a rubric in a valid and reliable manner when the same 52 pieces of student written work are rated by approximately 10 teacher participants.

Rationale

Some researchers (Hess et al., 2009; Thompson et al., 2008; Yassin et al., 2010) have already used BRT as a rubric; that is, they have used BRT to categorize student thinking as part of an assessment. A BRT assessment rubric could fill a gap in the local assessment system because there is no assessment of student thinking levels. Educators are not currently implementing a tool to score student thinking. The local learning community could benefit from the use of the BRT as a rubric, should it be found valid

and reliable. The rationale for this study is that rubrics must be shown to be valid and reliable before they are used to guide instructional practice (Brookhart & Chen, 2015).

This project study could contribute positively to the local setting and potentially far beyond because the teaching and assessment of HOTS are considered essential 21st century skills (Afandi, Sajidan, Muhammad, & Nunuk, 2018). HOTS are also time consuming to evaluate and require extensive amounts of student written work to track the development of student skills--for example, the large and growing stockpile of student written work in the school's new digital portfolio. In this study, I focused on evaluating the discriminant validity of BRT as a rubric by comparing previously assigned scores with the BRT rubric ratings. The primary purpose of this study was to ascertain whether teachers can score student writing with the BRT rubric demonstrating evidence of interrater reliability. I wanted to inform educators of the validity and reliability of using the BRT rubric to categorize student thinking on a continuum from low to high when evaluating written work.

Definition of Terms

21st century skills: The Partnership for 21st Century Skills Framework for 21st Century Learning identified four categories including (a) core subjects and 21st century themes, (b) learning and innovation, (c) information and media, and (d) technology skills and life and career skills (Voogt & Roblin, 2012).

Bloom's Revised Taxonomy (BRT): The original Bloom's Taxonomy consisted of a hierarchy of six cognitive processes: knowledge, comprehension, application, analysis, synthesis, and evaluation ((Anderson, Krathwohl, & Bloom, 2001). There is support in

the field for an adjustment of this hierarchy and the revision of the original highest category from synthesizing to creating (Amer, 2006; Radmehr & Drake, 2017; Wilson, 2013).

Digital portfolio: A web-based tool that allows students to develop, design, and manage project-based learning. This digital portfolio includes curriculum, planning tools, blogging, and a personal portfolio space for each student. Using a digital portfolio allows teachers to glean their students' creative processes, which allows them to coach students in a more personal way, according to the CEO of a portfolio software startup (2016).

Divergent thinking: The breakdown of a topic into varied components to stimulate creative thinking (Baer, 2014).

Formative assessment: Although many definitions of formative assessment exist, in this study formative assessment was the process of gathering the strongest possible evidence to document student learning to inform both students and educators to impact future instruction (Fisher & Frey, 2007; Wiliam, 2011).

Higher order thinking skills (HOTS): The higher degrees of thinking according to a cognitive taxonomy, as defined by Brookhart (2010). In a general sense, HOTS can be evaluated based on three different applications: transfer, critical thinking, and problem solving. Specific to Bloom's taxonomy, the three HOTS in the Amer (2006) revision are analyzing, evaluating, and creating.

Lower order thinking skills (LOTS): In Amer's (2006) revision of Bloom's taxonomy the three lower order thinking skills are remembering, understanding, and

applying. These are skills that current assessments capture but do not explicitly label as different levels of student thinking.

Problem-based learning: Student-centered learning opportunities in which students focus on an open-ended question or problem to which they propose a solution after following actual applications of content, skills, and the development of 21st-century skills (Vasan, Venkatachary, & Freebody, 2006).

Productive thinking: Mental activity that occurs when one combines knowledge with critical or creative thinking (Hurson, 2008).

Reliability: The degree to which an assessment tool produces stable and consistent results (Phelan & Wren, 2006).

Thinking: "Any mental activity that helps to formulate or solve a problem, make a decision, or fulfill a desire to understand. Thinking occurs when one is searching for answers, and reaching meaning" (Ngang, Nair, & Prachak, 2014, p.3760).

Validity: How well a test measures what it is intended to measure (Phelan & Wren, 2006).

Significance of the Study

If the teachers in this study can reliably and validly use the already existing and widely accepted BRT for the new purpose of a rubric with which teachers can score levels of student thinking in written work, then the school will have gained a new tool. For future use of the tool, educators should be trained to determine if a written answer falls in the BRT categories of remembering through evaluate. This simple step of using the BRT as a rubric to rate individual pieces of student work could potentially be

extended to other steps such as informing teachers about the need for remediation or being able to track developing student thinking across time. For example, a teacher could also determine if the student answers to a prompt the teacher wrote at the evaluate level were answered at the evaluation level of thinking; if not then teachers would know that student needed additional help with evaluation level HOTS. Teachers could also eventually use the BRT rubric across assignments in a unit to look for trends in student thinking levels. Indeed, because the BRT rubric is not content or grade specific, teachers could track the development of student thinking across the middle school education process.

The ability to track student thinking levels could help teachers design instruction that produces genuine learning. That is, scores at the higher levels such as evaluating and creating on the BRT scale denote that students are utilizing HOTS that are considered genuine. Indeed, using this BRT rubric for scoring student thinking-level progress means that educators would be able to document genuine learning. This genuine level of learning provides evidence of the capacity of a student to transfer learning to future real-life challenges. Additionally, with the BRT rubric educators could have more objective evidence to determine which instructional strategies they implemented had prompted the largest growth from LOTS to HOTS. In kind, they will also know which units need restructuring to promote genuine or HOTS learning. Each step along each student's individual learning path requires the consistent documentation and assessment of student thinking (Haynes, Lisic, Goltz, Stein, & Harris, 2016). If evaluated using the consistent, valid criteria of BRT, then educators can track student demonstrations of success along

the learning path from the BRT level remembering towards the highest BRT level of creating.

To systematically capture, store, and assess student work across time, new means of collecting and organizing student work is necessary. The charter school is an innovation partner with a pilot site for its digital portfolio program. This study is significant to the digital portfolio pilot site because the study is assessing writing samples produced within their system following their process for action-based learning. Furthermore, this partnership means that the school will receive full access to the latest product features. The school receives monthly site visits focused on customizing the product for students and teachers. It also means that the personnel of the digital portfolio wish to learn the findings of this project study and could potentially integrate BRT criteria into their software. If the short-term goal of showing evidence of discriminant validity and inter-rater reliability using BRT for scoring student work in the digital portfolio software were achieved, then there may be more software development that would allow additional educators to use BRT embedded within the software.

The long-term goal is to provide a window into the students' thinking processes, which will allow the teacher to coach students in a more individualized way. Opportunities are woven throughout this digital portfolio in which students apply what they have learned and work through the steps of solving problems. This learning process is a bi-functional process including both doing and thinking which capitalizes on students' level of cognitive readiness (Hung, 2006). Many business leaders say that the job of the future will be projects (CEO, personal communication, May 15, 2015). The

ability to design, manage, evaluate and collaborate on projects will be the key to success in almost any field. Increasingly, colleges are also reflecting this shift and accepting student portfolios in their application process.

The Massachusetts Institute of Technology has stated that the reason they are accepting portfolio-based applications is that they want to see how students respond in unstructured settings (Jaschik, 2015). By offering this increased support for project-based learning, this public charter middle school is on the leading edge of what many see as the future of education. If project-based learning can be evaluated using the BRT rubric, then the colleges will have even better information regarding the thinking levels of each applicant. Overall, it is important for assessment purposes that a student can demonstrate what and how they have learned and that they are capable of essential skills such as critical thinking, collaboration, flexibility, motivation, effective communication, assessing the relevancy of information, and curiosity and imagination (Wagner, 2014). The first step on this assessment journey was validly and reliably categorizing student-thinking levels using a BRT rubric on written work from project-based learning.

Research Questions and Hypotheses

I used a quantitative research design. The focus of the first research question was on the correlation between archived teacher grades for each piece of writing and the BRT scores for each piece of writing. The first research question pertains to discriminant validity--that is, do the BRT and teacher grades measure different things? The discriminant validity analysis was determined with a *t* test to assess whether there was a correlation between the teacher grades for each writing sample and the mean ratings

assigned by raters using the BRT. I expected that there would not be a correlation thus indicating that the teacher grades and the BRT ratings were distinct constructs. The focus of the second and third research questions was on examining reliability through two separate uses of the Krippendorff estimate. The second research question pertains to the Krippendorff estimate for the sample population. The third research question provided an estimate for the entire true population that the sample was taken from, and as such is an inferential statistic. The research questions and hypotheses were as follows:

RQ 1: What was the relationship between teacher grades for each writing sample and the BRT-based LOTS-HOTS ratings for each writing sample?

$H_0$1 There was no relationship between the classroom grades and BRT ratings assigned to each writing sample

$H_1$1: There was a relationship between the classroom grades and BRT ratings assigned to each writing sample.

RQ2. Was there a moderate (>.7) or better interrater reliability Krippendorf estimates demonstrated by middle school teachers' ratings using BRT for scoring multiple writing samples of student demonstrations of thinking?

$H_0$2: There was not a moderate (>.7) or better interrater reliability Krippendorf estimates demonstrated by middle school teachers' ratings using BRT for scoring multiple writing samples of student demonstrations of thinking.

$H_1$2: There was a moderate (>.7) or better interrater reliability Krippendorf estimate demonstrated by middle school teachers' ratings using BRT for scoring multiple samples of student demonstrations of thinking.

RQ3. Was there a moderate (>.7) or better interrater reliability demonstrated by the true population Krippendorff alpha estimates between middle school teachers' ratings using the BRT scoring rubric for multiple samples of student demonstrations of thinking?

$H_0$3: There was no moderate (>.7) or better interrater reliability demonstrated by the true population Krippendorff alpha estimates between middle school teachers' ratings using the BRT scoring rubric for multiple samples of student demonstrations of thinking.

$H_1$3: There was a moderate (>.7) or better interrater reliability demonstrated by the true population Krippendorff alpha estimate between middle school teachers' ratings using BRT scoring rubric for multiple samples of student demonstrations of thinking.

Review of the Literature

During this literature review, there were a series of topics that built upon one another. They are ordered from most basic to the next logical aspect of assessing thinking to consider. The first three headings of this review discuss the basic stepping stones of HOTS cognition including types of thinking, transfer, and motivation. The next section addresses the need for measuring both LOTS and HOTS in assessment and is titled: balanced educational objectives and tools for the 21st century. If assessments such as the BRT rubric do indeed identify the students' level of thinking, then it is necessary that teaching develop student thinking as is discussed in the section learning environments for demonstrating thinking. The last two sections address assessment starting with the prospect of assessing thinking and learning and then focusing more narrowly on learning portfolios as authentic assessment.

During the literature review, I focused on the terms *assessment*, *thinking*, *learning*, *learning progression*, *digital portfolio*, *Bloom's Revised Taxonomy*, and *writing*. Searches included the following indices and databases: ERIC, Sage, EBSCO, and ProQuest. To find additional research, I searched using terms associated with the *learning process*, and *assessment techniques.*

Conceptual Framework: Bloom's Taxonomy

Among the constructivist learning theories, there are social constructivist theories and cognitive constructivist theories (Biggs, 1996). Constructivist learning is an active learning process through which learners scaffold and adapt what they know according to new information (Shepard, 2000). Within constructivist learning theory there are two main assessment frameworks; a) authentic assessments which focus on higher order thinking and knowledge integration, and b) developmental assessments which focus on diagnosing a student's readiness in order to adjust instruction (Mokharti, Yellin, Bull, & Montgomery, 1996). This project study focuses on the first, authentic assessments. If the BRT rubric proves to be a valid and reliable for authentic assessments, teachers could then use it for developmental assessment purposes. Understanding the evolution and use of BRT is the foundation of this study.

BRT is based upon Benjamin Bloom's taxonomy, originally designed by Bloom in 1956 along with a group of educational psychologists, classified educational objectives into six categories (Sultana, 2010). After more than forty years of instructional design based on Bloom's original taxonomy, Anderson, Krathwohl, and Bloom (2001) revised the taxonomy to include the previously classified thinking skills as cognitive strategies in

verb form with create replacing evaluate at the top of the hierarchy. BRT is a widely-used guide for the design of curriculum and evaluation of instructional opportunities within the field (Forehand, 2010; Thompson & O' Loughlin, 2015). This project study will examine the use of BRT as a rubric of leveled categories for assessing thinking in students writing. Specifically, teachers will rate thinking in student writing as fitting into one of the six levels in Table 1: remembering, understanding, applying, analyzing, evaluating, and creating.

Table 1

Bloom's Revised Taxonomy Verbs Matched to Similar Verbs on the Higher Order Thinking Skills to Lower Order Thinking Skills Continuum

	HOTS
BRT verbs	Similar verbs
Creating	Designing, constructing, planning, producing, inventing, inventing, devising, making
Evaluating	Checking, hypothesizing, critiquing, experimenting, judging, testing, detecting
Analyzing	Comparing, organizing, deconstructing, attributing, outlining, finding, structuring, integrating
Applying	Implementing, carrying out, using, executing
Understanding	Interpreting, summarizing, inferring, paraphrasing, classifying, comparing, explaining, exemplifying

Remembering	Recognizing, listing, describing, identifying, retrieving, naming, locating, finding

LOTS

Note. BRT = Bloom's Revised Taxonomy; HOTS = higher order thinking skills; LOTS = lower order thinking skills. From "Bloom's Revised Taxonomy (BRT) Verbs and Similar Verbs from Higher Order Thinking Skills (HOTS) down to Lower Order Thinking Skills (LOTS)," by Crockett Global Citizen Staff Global Citizenship. Copyright 2017 by Global Citizen. Adapted with permission.

While changes have occurred in the approach to teaching, there is still a gap in the practice of developing and implementing assessments which require students to demonstrate higher order cognitive progressions including the BRT categories of evaluation and creation of new knowledge, as well as metacognitive awareness of these thinking skills (Draper, 2015). Bloom's Revised Taxonomy was developed to foster the development of assessments focused on varied cognitive demonstrations (Bezuidenhout & Alt, 2011; Rashid & Duys, 2015). Haolader, Avi and Foysol (2015) identify that this type of structured construction of knowledge occurs in the design phase of education. For example, BRT is used to design questions to ask students during small group discussions at a particular level such as the understanding level. Haolader et al., (2015) point out that BRT is rarely, if ever part of the design of assessment tools. This study seeks to use BRT explicitly for assessment as a rubric.

Indeed, most educators currently practicing in the field do not commonly assess BRT levels at any point. Instead, teachers' assessments largely focus on summative assessment of content recall and organization (Huitt, 2011). Educators could emphasize that instead of task completion, that the ultimate goal is profound and genuine learning.

Instead of just a grade, we could have an evaluation of whether thinking and learning had taken place (Brookhart, 2013). Teachers could have a consistent focus on student thinking assessment with BRT rubrics; teachers could use BRT rubric data as a central tool for driving the next instructional steps for all students (Wiliam, 2011).

Educators strive to stimulate higher levels of thinking through learning opportunities, therefore the assessment of student progress is required for continued growth (Cunningham & Lachapelle, 2014). To teach for advancement in student thinking, we must be able to assess student thinking as they develop their ability to apply content effectively (Nkhoma, Lam, Richardson, Kam, & Lau, 2016). The field requires the development of a quantitative assessment of thinking to track this growth and evaluate student preparedness to tackle tasks that require higher-order thinking (Rembach & Dison, 2016). The BRT rubric might be that assessment tool.

Types of Thinking

Thinking is constructed in a context. Much like instructional strategies vary based on the students in a given classroom, the type of thinking one employs depends on the application of thought required (Hung, 2006). Different types of thinking are good for different types of tasks, they are neither good or bad in their own right. In the event that a task requires divergent thinking, the thinker would generate as many possible solutions or theories as one can regard a concept or topic (Gallavan & Kottler, 2012; Kaufman, Lee, Baer, & Lee, 2007). Hurson (2008) described productive thinking as a process through which one combines knowledge with critical or creative thinking. My analysis of

literature reveals a gap in practice on assessing students for the critical capacity of either divergent or productive thinking (Lam, 2017).

This gap is highlighted by the historical emphasis in curriculum and standardized assessments on convergent thinking tasks, or tasks in which thinkers are expected to apply content or knowledge to complete a finite or defined task. There is a lack of instruction in using converged ideas or content associations to create diverging solutions to proposed challenges (Kaufman et al., 2007). Recent attempts have been made to infiltrate standardized assessments with performance tasks which require varied levels of divergent thinking (Common Core State Standards Initiative, 2013). There appears to be a disconnect between the convergent thinking required on tests and the divergent thinking required for solving real-world problems.

Tests and real-world problems both have objectives. Governments and schools list objectives in standards, and then assess based upon those standards. While students are completing tasks in school, the idea is that they learn and develop an understanding of a concept or skill often tied to a standard or benchmark. However, a student may arrive at an answer being unsure of how they got to the answer, because subconscious connections were being made by their mind all the while they were working on a task (Runco, 2014). That is, we rarely can see a person's thinking but rather simply a finished product that reflects the scaffolded thinking used to create the product (Sotiriadou & Hill, 2015). It may be useful to have assessments of varied levels and applications of thinking. Assessment tools for evaluation of applied thinking and transfer of knowledge would provide a gauge for educators, and be useful in the practice of prompting learning

(Harvey & Daniels, 2009; Kleickmann, Richter, Kunter, Elsner, Besser, Krauss, & Baumert, 2013; Pascal, Tíjaro-Rojas, Oyander, & Arce, 2017).

Runco (2014) demonstrated that it is through subconscious associations that learners shift their level of understanding, while thinking, from superficial representations (content knowledge) to complex representations and transfer. Once one has reached the more complex levels of thinking, genuine and lasting learning has occurred and independent transfer is possible in new and unknown situations (Dagostino, Carifio, Bauer, Zhao, & Hashim, 2015). In this same vein of learning Argyris and Schon (1974) identified single and double loop learning as components of their theory of action in which human beings are agents of change. Single loop learning identifies one's decision to follow existing rules, while double loop learning (representative of middle levels of thinking in BRT) occurs when one adapts their thinking and generates ideas about the existing rules (Argyris & Schon, 1974). Triple loop learning takes thinking to the highest level of BRT (create) and occurs when on creates new rules based on what they have learned about a certain topic or situation (McNamara, 2006). This notion of transfer, of taking knowledge and applying it, is important because it is the ultimate assessment; do students use what they learn in situations outside of classrooms?

Transfer

Transfer occurs when prior learning influences future performance (Clark, 2011). Varied levels of transfer have been noted: near, far, and further transfer. The degree is based on the connection and similarities between the knowledge and the situation in which one is trying to perform a task that requires that knowledge (Kaiser, Kaminski, &

Foley, 2013). Brent (2011) asserted that after the transfer of knowledge has occurred, the new resulting knowledge has been transformed – the knowledge is now associated with the situation in which is successfully helped solve the problem. A classroom focused on problem-based learning offers potential to observe stages of knowledge incorporation, transfer, and transformation through various instructional strategies (Panasan & Nuangchalerm, 2010; Tidwell, 2015). Having a learning environment and educator designing opportunities for transfer is helpful.

The Experiential Learning Theory designed by Kolb (1984) identified a four-cycle learning process in which once associates concrete-abstract and reflective-active dimensions of learning. This cycle of learning begins with an experience, followed by an assimilation of the new knowledge with old values to be reflected on and transferred from abstract thoughts to concrete associations (Kolb, 1984). To further understand the critical nature of transfer as an ultimate test for learning, we can examine the biological aspects of the physical learning process. Zull (2011), drawing from the prior works of Kolb's Experiential Learning Theory, proposes that the brain physically changes as one learns though the process of what he call the Four Pillars: Gathering Information, Reflection, Creating, and Testing. Throughout this process Zull (2006) noted that in the early phases of learning one gathers data through sensory inputs and assigns a value to each gathered data point. The process through which the data moves from the sensory neocortex to the association regions Zull (2006) labels as the reflection phase. This is followed by the creation phase in which these new associations engage working memory to create new ideas or theories. The final pillar of testing engages the motor brain to transfer of the

created theory from abstract to concrete through application to a new situation or challenge (Zull, 2006).

Indeed, according to Nokes (2009), there is a need for educators to provide learning experiences that include the meta-cognitive practice of transfer. Students must be explicitly taught how to become cognizant of opportunities in which they may transfer knowledge and have the skills to proceed. Ultimately, steps in the instructional process are needed during which transfer skills are explicitly taught and transfer itself is measured (Nokes, 2009). Research in the field emphasizes the importance of educational opportunities focused on the transformation of knowledge which foster growth in citizenship and the development of social involvement (Gardner, 2010; Gerlach & Reinagel, 2016). For students to reach their full potential in terms of transfer, research shows that they should be intrinsically motivated and acting on their volition. It is not enough to have instruction; students' emotional state must be figured into the learning equation (Zull, 2006).

Motivation, Volition, and Engagement

Consistent student engagement and achievement of long-term goals requires the existence of motivation and volition (DeBarger, Dornsife, Rosier, Shechtman, & Yarnall, 2013). Jones (2012) recognized the need for educators to provide relevant, real-world learning activities including problem-solving, critical thinking, and engagement to foster the development of twenty-first-century skills. Student-engaged assessment as a framework for evaluation provides an opportunity for students to investigate their own growth and capacity building through self-directed learning (Berger, Rugen, & Woodfin,

2014). It could be motivating for students to self-evaluate their work possibly using the BRT rubric (Hammill, Best, & Anderson, 2015). Zull (2006) proposed that educators who are aware of the plasticity and physical changes in the brain when learning occurs are better suited to design experiences and instructional opportunities that will link emotion to thought in an intentional way yielding a more engaged and motivated student. Educator's awareness of the interconnectedness and links throughout the brains framework and how they influence thinking and engagement will increase the likelihood of the educators reaching the student to engage them in HOTS (Siegel, 2010).

Different people have conceptualized how to motivate students in classrooms. Pearlman (2010) identified effective twenty-first-century schools as those in which there are students at work. Long (2012) asserted that students should be empowered to thrive through participation in Design Thinking. Design Thinking in itself is engaging, because it focuses on the "improvement of the human experience through educational opportunities that combine ongoing collaboration, systematic thinking balanced with creativity and analysis" (p.14). For example, students may do the work of science experiments on a local river to inform water specialists about the health of the water as well as design and implement water improvement projects themselves. Bezuidenhout and Alt (2011) noted that students must be engaged and see value in tasks they are completing at any point in the learning progression for lasting change to occur which will yield deep and significant learning. In order to have students go through levels toward the HOTS in BRT, they must be engaged in meaningful work with transfer opportunities (Gilboy, Heinerichs, & Pazzaglia, 2015).

Balanced Educational Objectives and Tools for the 21st Century

For lasting changes to occur in education, it is imperative that policymakers, administrators, and most importantly practitioners recognize necessary changes in learner expectations as well as the purpose of teaching; teaching students to think (Retna & Ng, 2016). The initial shift requires the transition from teacher as keeper of knowledge to the teacher in the role of facilitator and guide (Shepard, 2000; Dolan & Collins, 2015). Collet (2014) emphasized a balance between self-direction and expert mentoring as the key to successful learning. A key component to fostering genuine learning is learner participation with a classroom dynamic rooted in the value of developing thinking skills. Students in a reflexive and thought-based classroom are likely to own their learning processes, and emphasize questioning as a method of learning (Peen & Arshad, 2014). Student development of questioning techniques provides a method for motivating and engaging students in authentic concerns that they may have or passions they chose to pursue while promoting collaborative dialogue and other necessary 21st century skills (Rothstein & Santana, 2011). Beyond focusing on content as the only objective, the development of quality thinking is a higher educational objective and goal (Choudhury, Gouldsborough, & Shaw, 2015).

To engage students in 21st-century habits of learning content and thinking, the design and implementation of a problem-based learning program offers a combination of the elements more supportive than traditional spoon-feeding of information. Within science classrooms, the heuristic inquiry approach is used to learning concepts and skills within the domain (Günel, Memis, & Büyükkasap, 2010; Lo et al., 2016). A heuristic

learned through discovery or inquiry provides improved understanding, increased connections, and an increase in cognitive activity (Al-Fayez & Jubran, 2012). In short, heuristics are common ways of thinking that can be applied, or transferred into new situations.

Therefore, we need learning environments in which an educator designs opportunities for students to engage in the active discovery of methods and heuristics of thinking (Bezuidenhout & Alt, 2011). This type of learning design requires the use of processes and instruments (such as the BRT rubric) for gauging thinking and student growth towards independent near transfer (Anderson, Krathwohl, & Bloom, 2001). Hong and Choi (2011) examined the relationships and patterns occurring during reflective thinking of novice to expertise in a field. Hong and Choi (2011) was working on developing a research-based learning progression that students travel from novice to expert. The BRT rubric is a more general progression from the novice level of remembering to the most expert level of creating.

Learning Environments for Demonstrating Thinking

Thinking is an internal process that we cannot see, so we must depend on models and research documenting best practices to encourage students to develop and practice higher quality thinking. In the process of learning, students filter through their personal knowledge base, experiences, and internal reactions. Through sound instructional practice students process new associations and genuine, transferable learning occurs (Spruce & Bol, 2015). Ultimately, based on the learned ability to think well, good thinkers develop original ideas and thoughts to help them solve future challenges (Kahneman & Egan,

2011). Students must demonstrate thinking mastery on assessments and other learning tasks regularly as designed, implemented, and monitored by a skilled educator (Tíjaro-Rojas, Arce-Trigatti, Pascal, & Arce, 2016).

Practitioners should assess students' process frequently and regularly. The gleaned data should be used to influence future instruction. Documented evidence of students thinking as they progress towards learning should occur in a formative, ongoing thread focused on providing a gauge for educators and students through a multifaceted reflective cycle (Brookhart, 2013). The path of learning winds in many directions, and assessment is a necessary feedback loop to stay on the course toward the instructional objective. Collecting, organizing, and maintaining the scoring of writing assessments has been a long-standing challenge, but with new technology there are new opportunities as discussed below (Conley, 2015).

A project-based learning portfolio approach provides students with a software platform to complete activities and associated writing samples at various phases throughout a project timeline, thus providing necessary evidence of student thinking and learning. A digital portfolio system organizationally supports tracking and assessment of students' development of thinking ability thus fostering the creation of new knowledge out of existing information (Fink, 2003). Educators can then score students' work using a cognitive progression such as BRT to provide objective, structured feedback to track thinking through the varied demonstrations collected in a digital portfolio system.

Assessment of Thinking and Learning

As students grow and develop academic skills, their teachers must be proficient in assessing if their students are growing along the way (Black, Harrison, Lee, Marshall, & Wiliam, 2003). Formative assessment provides a pathway along which educators can evaluate if students are learning. Focusing on which areas students are struggling with or have mastered to adjust instruction thus ensuring the next educational opportunity the student experiences yields evidence of learning (Black & Wiliam, 1998). Often the breakdown when using formative assessment occurs when the educator gathers rich data on a student and then fails to make changes that impact instruction in the future (Black et al., 2003). Formative assessment is one important piece of the assessment package.

In a society with numerous factors influence students' learning, educators must be proficient in controlling the one factor they have access to, the efficacy of instruction they provide each student. Formative assessment is one of many success indicators that provide a glimpse into the learning necessary for a student to independently transfer content and skills to new, similar situations (Hargreaves, 2003; Hernández & Rodríguez, 2016). Assessments, however, are not useful without quality rubrics that track students' ability to move through cognitive levels to reach higher-order levels such as evaluation, creation, and self-awareness (CEO, personal communication, November 5, 2015; Young, James, & Noy). Scoring rubrics which prompt metacognitive evaluation should be provided to students on a daily basis in a learning environment where the established climate permits thinking and learning to occur (Brookhart, 2010).

Positive classroom environments explicitly engage in teaching metacognitive practices including retrospective self-reporting (Sabourin, Lowe, & Bowman, 2015). These metacognitive practices are critical to the development of productive thinking and student progress in a domain (Gilmore & Feldon, 2010). Gilmore and Feldon (2010) further discussed the levels of self-reporting and metacognitive practice along an expertise continuum. That is, students change how they self-report their learning as their expertise grows; this is described next. This continuum begins with novice as a very fact based reproductive thinking (mimetic) level. This is similar to the remembering level in the BRT rubric. Working towards an intermediate standard of thinking would be possessing procedural schema. This includes the capability to recall and filter a large pool of knowledge. Finally, the expert, possessing increased ability to filter information using working memory in an automated manner, provides the space for divergent breakthroughs. Yoruk and Runco (2014) found that at the expert level, there is an inability to recall the smaller steps leading to the finished product due to automaticity and the ability to make subconscious connections. This general progression from declarative, procedural, and conceptual understanding is found across domains, and is similar to BRT in that way. In all domains, the issue that remains challenging to researchers is to note patterns between discrete elements of thought and universal intellectual standards (Lai, 2011). This is beyond the scope of this project study.

Authentic Assessment of Learning Portfolio

Based on a nationwide call to action requiring a renewed knowledge paradigm of HOTS goals, educators must begin to develop authentic assessments (McTighe &

Wiggins, 2013). An authentic assessment requires the transfer knowledge to a real-world application (Mueller, 2016). Kleickmann et al. (2013) emphasized the rethinking and integration of authentic assessments focused on the higher-order skills; creativity, collaboration, and filtering. These should be measured throughout the learning process via formative assessment.

Digital portfolio documentation is a key method in making students' internal thinking and learning visible to assessors. Bjornavold (2009) validated the digital portfolio, as a method for collecting evidence of authentic assessment. Students' development of their portfolios is a dynamic assessment practice that addresses twenty-first-century learning and characteristics of a renewed knowledge paradigm (Besser, 2011). The written component of a digital portfolio, provides evidence necessary to the assessment of thinking at varied levels, but only if there is a valid and reliable way to score the thinking such as the BRT rubric in this project study.

The inclusion of writing in a digital portfolio provides a structure for monitoring students' development of thinking ability as well as the key feature to monitor learning over time electronically (O'Brien-Moran & Soiferman, 2010; Wason, Sinvhal, & Bhattacharya, 2016). The written work to be evaluated in this study will come from a digital portfolio and multiple examples of student written work will be evaluated.

Under the framework of social constructivism, this review has discussed general ideas about thinking, motivation, and assessment. These are grounded in ideas that learners construct understanding through different types of experiences focused on thinking and this construction is based on intrinsic motivation in the best of cases (Kolb,

1984; Zull, 2006). The implications of this literature review are that teachers should acknowledge and focus their attention on the internal processes of thinking that their students are developing (Gilmore & Feldon, 2010).

Implications

While there are different categorizations of types of thinking, BRT has persisted through time and takes a central role in current reform teaching practices. These six BRT levels are a successful way of framing thinking. The implications of this for this study is that BRT can be used in an attempt to categorize student written work as demonstrating one of the six BRT types of thinking. In order to study thinking, learners must experience learning environments that elicit different levels of thinking and that use formative assessment to track that thinking (Brookhart, 2010).

In this study, I will use one of the recommended authentic assessments. It is a digital portfolio that contains all of the written work a student has completed across an authentic problem-based learning unit (Bjornavold, 2009). For this study, I will test the validity and reliability of using the BRT rubric. If the BRT rubric is valid and reliable, then it may be used to score student work across time and the different levels of prompts during the problem based learning process (Bauer, 2016). Ultimately, this study could provide reliable and valid scoring of varied levels of students thinking using the BRT. Based on my research findings I will develop and include an appropriate application project in Appendix A.

The data from this project study may indicate that scoring writing samples for varied levels of thinking using the BRT is valid and reliable across raters. If the data

trends in this direction after collection and analysis, one possible direction that the project study would take is the development of professional development workshops focused on scoring student thinking using the BRT. If it does not, then a policy recommendation for additional testing of the BRT and recommendations for other possible tools to evaluate student thinking may be developed for the charter school administration and teachers.

Prior to the designing of the professional development workshops, the author would develop a handbook or manual for teachers focused on the assessment of student thinking. Components of this handbook/manual on assessing thinking would include an overview of the levels of thinking including criteria and reference verbs for identification of varied levels, exemplars of student writing samples demonstrating the highest levels of thinking, and steps for scoring writing samples reliably between raters. Once the data is analyzed, there may be additional components of the handbook/manual that would need to be included. Such a handbook could be another direction for the project proper.

Once the handbook for assessing student thinking was developed and approved for implementation, the author could begin sessions during which faculty are trained in each component of the handbook. Upon completing of training in the process of assessing student thinking using the BRT, sessions could occur during which educators reflect on how their practice has changed based on their capacity to reliably score student thinking. In the future, educators could use this handbook and scoring process as a stepping point for further study of how student thinking capacity if growing over time to inform their practice and instruction.

A final important implication of this study is that students need thinking skills to get desirable jobs. Rather than simply needing to regurgitate content when prompted, viable candidates for a job or project must be able to locate, filter, select, apply and manipulate content to fit a solution they propose. Ultimately, expert thinkers can organize and simplify their explanations (Dowd, Duncan, & Reynolds, 2015). Business leaders around the globe are noticing that the biggest challenge they have lies in finding graduates prepared to take a project from its start to its finish without requiring consistent direction (CEO, personal communication, June 10, 2015). They desire employees who can think. Using BRT as a rubric may help teachers plan and monitor student thinking level abilities. Positive social change is achieved when students develop and apply higher order thinking skills for work and life.

Summary

This section began with a discussion of challenges facing educators while teaching students to think. This is in response to indicators that students graduating high school are ill prepared to become a contributing member of our global knowledge economy. Specifically noted are the gaps in teaching practice around the documentation and evaluation of students thinking. Also examined was the problem of simply capturing significant learning without assessment. Additionally, included in this section is a description of 21st-century objectives and the misalignment between what content students are expected to be proficient in versus what thinking skills are emphasized in research and business. The remaining section of this project study discusses the

methodology and plan used to collect and analyze data, as well as protect the rights of study participants.

Section 2: The Methodology

Introduction

The purpose of this project study was to examine the discriminant validity and interrater reliability of BRT as a rubric for scoring students' writing and measuring the progression of student levels of thinking. The goal was to develop a reliable and valid method to objectively score students' thinking levels through written work. To investigate the discriminant validity of the BRT rubric, I compared the ratings assigned to each writing to the grades teachers had assigned to determine if they were correlated. Examining the discriminant validity in this project study entailed determining whether the ratings assigned to each writing sample were unrelated to the grades previously assigned to the writing samples. If they were not, and if the interrater reliabilities were sufficiently high, I then concluded that the classroom teacher and teacher using the BRT rubric were grading two different constructs: content knowledge for teacher ratings and student thinking for the BRT assessments. I calculated the interrater reliability between educators for scoring student writing using the BRT rubric.

For my study, I focused on the assessment of evidence demonstrating students' developing thinking capacity from the lower levels of thinking to the higher levels of thinking. In this case, I used archival data based on the published writing pieces that were pulled from the digital portfolio interface. The authenticity of the writing samples provided a view of potential daily use of the BRT rubric by allowing for a retrospective evaluation of the archived writing samples.

Research Design and Approach

I used a quantitative, nonexperimental research design to investigate the discriminant validity of the BRT rubric and the interrater reliability between teachers scoring samples of student writing with the BRT rubric. A nonexperimental study includes the study of variables and not the manipulation of variables within the existing context (Creswell, 2014). My research design is an approach in which data collected during the study were analyzed including the nonmanipulated variables contained in archived writing samples. I evaluated the variables where they were in the context in which they occur naturally in the writing process. The independent variable in the study was the student writing samples. The dependent variables were the teachers' ratings. Interrater reliability refers to the degree to which two raters agree in their determination of a score or judgment (Phelan & Wren, 2006). As the scoring of writing is considered relatively subjective, the investigation of interrater reliability in scoring writing samples using the BRT rubric could be useful to the field because it may provide reliable criteria for quantifying students' ability to think at higher levels as demonstrated in writing.

Initially, I considered different qualitative designs such as conducting a grounded theory-based case study to develop a theory inductively based on the current assessment of students' thinking. Because the school and teachers were lacking a tool for evaluating thinking, my focus turned to locating a way to reliably evaluate student thinking. With this in mind, I focused on the purpose of this study (i.e., my aim to contribute to the practice of assessing and evaluating thinking using a valid, reliable scoring structure).

I then considered conducting a phenomenological case study documenting student thinking to then generate scoring schema used to analyze and code levels of thinking (see Lodico, Spaulding, & Voegtle, 2010). Additionally, I considered conducting a narrative analysis case study to gather data form from the perspectives of students using their own voices. However, the case study format would have constrained the study to a smaller sample of student work that would likely have been content dependent and therefore less useful for the broader population of teachers and students. A narrative case study would not have lent itself to the documentation of students' thinking as captured and documented through writing. A narrative case study would also neglect the broader quantitative evaluation of a scoring rubric based higher-order thinking schema (Rembach & Dison, 2016)..

Finally, I also considered a descriptive case study. Case study researchers follow a process of intensive analysis of a particular event within a bounded system to create a detailed understanding of that event (Creswell, 2014). In this case, the event was student thinking at a small charter school. In the study I focused on uncovering levels of thinking in student work and categorizing students' cognitive readiness and capability for higher order thinking. Thus, the entire focus was on students' cognitive abilities and the increasing use of higher levels of thinking (see Abrami et al., 2015). The problem was the focus on the evaluation of student thinking rather than content memory. In addition, the school leadership desired a measurement tool that could be used across classrooms as a general measure of student thinking rather than a content-dependent measure. The goal to develop and validate such a rubric clearly indicated the need for a quantitative approach.

Therefore, in this study I conducted a quantitative analysis of how teachers used BRT as assessment rubric criteria to evaluate students' levels of thinking using archived captures of student writing. In essence, in this study I attempted to quantify students' levels of thinking based on their writing using BRT. Through this study I assessed the practice and reliability of scoring students' writing with quantifiable BRT as a rubric code. The question was whether this tool would reliably identify students' levels of thinking across different teachers' evaluations of the same student work.

Setting and Sample

All samples for this quantitative study came from within a bounded system: one small, public, rural charter school, serving 290-300 students. I recruited teacher participants from two public, charter schools each serving 300 students. All teacher participants were recruited by responding to an e-mail invitation to participate in a study to validate a new rubric based on BRT for scoring levels of student thinking from samples of actual student essays. The teachers selected for participation in the interrater reliability section of this study were chosen using purposive sampling from the identified population to build a sample from which I was able to derive statistical inferences (see Etikan, Musa, & Alkassim, 2016). The teachers who completed the rating for the interrater reliability ranged in experience from one year of classroom teaching experience to 15 years of classroom teaching experience. The sample consisted of eight teachers, each scoring two pieces of published writing collected from 26 seventh-grade students. The purposive sample of teachers work at a charter school that promotes the development of HOTS.

I selected a purposeful sampling of two writing samples from each seventh-grade student's written work. All writing samples were collected from students utilizing the digital portfolio process. The deidentified writing samples were obtained from regularly assigned student work. For each deidentified writing sample, I only received the teacher's grade for that portfolio submission to utilize for the *t* test to evaluate discriminant validity. I chose seventh grade because it represents the middle of middle school and had enough students to supply sufficient number of samples for coding ($N = 52$). I estimated that participating teachers would take 5 minutes to rate each piece of written work for approximately four hours. Although this was a quantitative research design, purposeful sampling was required because I attempted to determine the interrater reliability between educators within a bounded system. The process for determining interrater reliability was defined by teachers who used a scoring rubric to assess thinking based on samples of seventh-grade student writing.

The Krippendorff estimates used in this study are point estimates with an inferential statistic regarding the full population true score. Krippendorff estimates do not require a power analysis to determine the number of raters or samples of work being rated (De Swert, 2012). The example Hayes and Krippendorff (2007) provided only had two raters with three samples of work each to demonstrate the power of the estimate. To ensure valid results, I used a minimum of eight raters and 52 pieces of published writing, far exceeding the minimums set forth by Krippendorff (see DeSwert, 2012).

Considering the participation of a greater number of raters, I reduced the number of writing samples from the referenced example of three pieces of work to two pieces of

work, which allowed for a reasonable amount of time to be spent scoring the pieces of writing per rater. Increasing the number of raters increased the number of writing samples to be scored in the allotted time frame, thus increasing the power of the estimate (Meyvis, van Osselaer, & Stijn, 2018). The number of scored pieces of writing with two per rater is still larger in number than if five raters scored three pieces each.

Instrumentation and Materials

Instruments

The instrument used to score writing in this study was the BRT (including a list of verbs for each level) as a rubric found in the literature (Crokett, 2018). The BRT was created to help organize levels of thinking and is used as a guide for generating classroom assignment prompts and assessment questions that ask for different levels of thinking. This study was different because it used the BRT to categorize student's written responses to prompts. Other researchers (Yassin et al., 2010; Amer, 2006; Hess et al., 2009; Thompson et al., 2008) have used BRT as a rubric, but did not evaluate the BRT rubric for validity or reliability. My study filled a gap in practice by evaluating the BRT as a valid and reliable process for assessing LOTS-HOTS. I provided the participant teachers with copies of the BRT as a rubric for evaluating student thinking using Bloom's Revised Taxonomy, together with a list of verbs associated with each rubric level (Heick, 2016). It was my hope that the list of verbs strengthened the BRT as an evaluation process by enhancing its reliability and validity. To facilitate the process of rating, teachers entered ratings into a scoring sheet using GoogleSheets®, which is exemplified in Table 2.

Table 2

Bloom's Revised Scoring Sheet

Student Work ID#	BRT level score
Sample 1	
Sample 2	

The six levels of thinking constructs included on the BRT are (a) Remembering, (b) Understanding, (c) Applying, (d) Analyzing, (e) Evaluating, and (f) Creating. These were numbered on the rubric from lowest (1) to highest (6), as assigned by the teachers. Thus, all level 2 responses were considered as the understanding level of thinking. A mean score of 2.1 – 2.9 was interpreted to represent the understanding level because the range was clearly situated between 1.1 – 1.9 (Remembering) and 3.1 – 3.9 (Applying). For each sample of writing, all teachers' ratings were averaged and a standard deviation was calculated to provide the descriptive statistics. Krippendorf estimates were calculated using the raw data with the KALPHA macro within the statistical software SPSS. Through this process, I generated a KALPHA discriminant validity estimate and reliability estimates for using the BRT as a rubric, a process that has not previously been accomplished. The first research question guiding this study examined the correlation between archived teacher grades for each piece of writing and the BRT scores for each piece of writing. The first research question pertains to the discriminant validity; do the BRT and teacher grades measure different things? The discriminant validity analysis will be determining with a *t* test if there is a correlation between the teacher grades for each

writing sample with the mean ratings assigned by raters using the BRT. It is expected that there would not be a correlation thus indicating that the teacher grades and the BRT ratings were distinct constructs. Where the second and third research questions examined reliability through two separate uses of the Krippendorff estimate. The second research question is the Krippendorff estimate for the sample population. The third research question provides an estimate for the entire true population that the sample was taken from, and as such is an inferential statistic.

To assist participants with their ratings, a list of 249 verbs (Appendix C) were shared with the teacher participants. This verb list was shared on a on a single sheet of paper, front and back. The paper of verbs and the paper of the rubric were the only two sheets of paper the teachers will use to rate the student work in a GoogleSheets®.

Materials

To select writing samples, I identified prompts within existing problem-based learning units that were likely to prompt a range of thinking. For example, a prompt that elicits only the first BRT level remembering is highly unlikely to have students writing at the fourth BRT level analysis. Second, I chose prompts from varied points of the problem-based learning process. The prompts from the late parts of the unit were intended to elicit BRT levels 5 evaluate or 6 create. For example, students were prompted with activities that asked them to collaborate, which led to actions occurring in the 'create' level of BRT. These selected prompts can be seen in Table 3.

Table 3

Bloom's Levels of Understanding and Question Examples

Level of Prompt	Prompt
Knowledge	What is your idea? Briefly describe what you will do. Is it clear? Is it specific?
Apply/Analyze	What else do you want to learn? How do you want to grow personally? What communication or technical skills do you want to gain?
Evaluate/Analyze/Apply	Impact: Does your idea help someone? Does it change or improve something? Does it allow you to learn something?
Create	Showcase: What will people see when they experience your solution? What will people experience at your showcase? What is your portfolio message?

Teachers who scored the data were not the seventh-grade teacher of the students whose archived work was chosen for the study. Teachers from this school were not responsible for actively trying to teach the BRT levels of thinking beyond general instructional best practices. All data to be scored was archived in a digital portfolio system. The written work was printed and unidentified for the purpose of scoring in this study to affirm confidentiality. Table 3 gives some example of the types of prompts used to elicit written responses that were scored using the BRT as a scoring rubric.

Training Process

To ensure that participants understand the scoring processes, I provided a training session during which participants were provided an opportunity to sort and score similar

writing pieces using the BRT as a scoring rubric (see Appendix E). The training was intended to share an overview with the participants of the BRT levels of thinking and a brief overview of the scoring template. The training provided participants with two sorting sessions as well as a warm-up scoring session prior to evaluating the actual writing samples for the study. Training was conducted for approximately 60 minutes. The training session closure included a 10-minute check for understanding during which each participant was given the opportunity to ask questions to clear up any confusion regarding the scoring process. The BRT (including the verb list for each level) as a rubric and the paper with the BRT verbs used in the training were the same as used in the study. The second portion of training included a warmup for participants to score similar but unrelated writing samples using the BRT. These samples were selected from the same grade, were the same length, and prompt type as those in the study, but were from a different assignment. During this training session, all participants were given the opportunity to ask questions regarding scoring writing samples using the BRT levels as the rubric. All participants were present during the training and the warm up. All participants had equal access to the same materials, warm up samples, and materials.

Data Collection and Analysis

Participant teachers utilized a GoogleSheet® (Table 2), shared with each participant in Google Drive, that acts as the confidential recording medium for the scoring process. The GoogleSheet® auto generated responses confidentially as designated in the form creation to not collect or record the user, in this case the

participant. All participant scoring responses were confidentially generated and only associated with randomly assigned rater identification numbers.

The process involved participants using one sheet with BRT scoring key and one sheet including 249 action verbs drawn from Bloom's Revised Taxonomy (see Appendix B). Each of the 8 participants had their own computers with GoogleSheet®. They were organized around one room, using privacy screens, so that that were not able to see each other's ratings. It was imperative that scorers be unaware of others scores to ensure that we can determine the inter-rater reliability of the BRT as a rubric. The teacher participants had confidential participant codes and coded each capture in the same order of presentation from earliest to the latest. These participant codes allowed for the organization and management of the data in a confidential manner.

The actual samples of written work were hardcopies numbered in order with 5 digit codes to increase anonymity. I personally accessed the existing student work directly from the digital portfolio and print paper copies. I removed any identifying marks as needed. The 2 samples per student were taken from one 7th grade class. No one except the teacher, myself, and the executive principal knows the identity of this teacher. This 7th grade teacher was not be a volunteer for the study. It would have been ideal to use the digital portfolio system but it would have been challenging to hide the identity of the students. To protect the student identity, numbered paper copies were supplied to teachers. Each participant had a total of 2 captures for each of the 26 unidentified students for a total of 52 written samples to score. Educators were given as much time as they needed to score all samples and record their scores on the spreadsheet. The expected

amount of time was 5 minutes per writing sample for a total of 260 minutes or approximately 4 hours and 15 minutes. There were snacks and a lunch at the end of coding. Teachers were be instructed to take breaks as needed.

Data Analysis Plan

Descriptive statistics. Once all scoring was completed, and all captures scores were recorded in the spreadsheets, the inter-rater reliability between the 8 participants was evaluated. This analysis used descriptive statistics to identify the mean and mode ratings for each of the 52 instances of student work that was scored by the teachers. These data were sorted by the mean score from lowest to highest to present a view of how many student captures tended to be rated highly, moderately, and low. Standard deviations were reported for each piece of student work to give a sense of how varied the ratings were for each student. The mode statistic indicated what rating was applied most often by the teacher participants. An example of the descriptive statistics table I planned to use to capture and display these data is provided in Table 4.

Table 4

Student Captures Ordered by Mean Score from Lowest to Highest

	Mean	Mode	Standard Deviation
Student 12345			
Student 23456			
Student 34567			

In addition to providing data on each instance of student work, I provided descriptive statistics for each of the teacher's overall ratings across all student work. This will reveal any bias teachers might have. An example of how this process was planned to work is provided in Table 5. The mean rating given by Teacher 1 across all student work in the example below is 2.3 compared to the mean rating for Teacher 2 of 4.5. These two teachers could be said to be typically different than one another in ratings. In terms of their modes; Teacher 1 applying the rating of 2 most often, and Teacher 2 giving the rating of 5 most often, again emphasizing their differences. Finally, the standard deviation of ratings applied by Teacher 1 was only 1.2 meaning that she did not have a wide range of scores. In contrast, Teacher 2's rating produced a standard deviation of 3.9 indicating that this teacher applied a wider range of ratings than Teacher 1 (*SD* = 1.2).

Table 5

Teacher Ratings Across all Student Captures

	Teacher 1	Teacher 2	Teacher *n*
Mean	2.3	4.5	
Mode	2	5	
SD	1.2	3.9	

Finally, I also used the mean scores to report how the two captures from each student were rated by the teachers. This process helped to reveal scoring trends in the same student's work. An example of the mean ratings table is provided in Table 6.

Table 6

Mean Ratings on the Two Samples for Each Student

Student #	Sample 1 mean	Sample 2 mean
Student 12345		
Student 23456		
Student 34567		

Validity

Often times it is useful to establish convergent validity for a measure by using two different research methods to determine if they both are measuring the same construct, thereby providing evidence that the construct itself exists (Trochinm, 2006; Rojas & Widiger, 2014). Given that the BRT has been tested for more than 20 years for its ability to categorize levels of thinking, it was deemed that convergent validity would already have a high likelihood of existing. In addition, adding another data collection method is beyond the scope of this project study. In contrast, determining discriminant validity is of great importance because if the BRT rubric is not assessing something other than what

the teachers' grades are already capturing, then this decreases the need for a separate method of evaluating student thinking. The intention is to determine if the students' grades for the content knowledge in their writing are correlated with the teachers' BRT ratings. The grades will already be established and collected from the teacher for the confidential identification codes. The ratings will be collected in this study. The two sets of ratings will be compared in SPSS using a Spearman rank correlation because the BRT ratings are categorical data (McDonald, 2009).

Reliability

This method of data collection and analysis lends itself to the use of inferential statistics as the study aims to rate the reliability of 8-10 raters using the same scoring rubric on the same student samples (Hayes & Krippendorff, 2007). Using inferential statistical analysis allowed the author to draw inferences around the population regarding the reliability of teachers using the scoring rubric (Angell, 2015). The Krippendorff Inter-rater Reliability Estimate was employed as the inferential statistical analysis to determine the instrument's reliability. In itself, the statistic is not inferential because it is a point estimate of the inter-rater reliability.

The use of the Macro KALPHA in SPSS, however, does produce inferential statistics related to the Krippendorff (Hayes & Krippendorff, 2007). Specifically, the Macro KALPHA in SPSS uses a bootstrapping method on the collected data to give an estimate measure of the true population alpha. That is, it allowed the inference of the true alpha of the larger population, from which the participants and captures were taken, and from which the inter-rater reliability for the entire local population of teachers and

students was calculated. KALPHA also reports the probability that the true alpha would lie below different minimum thresholds. For example, from sample analysis there may be a 3.23% probability that the KALPHA would be less than .8 for a population (De Swert, 2012). This would indicate a high probability (96.77%) of a good inter-rater reliability (KALPHA > .8) for both the sample and the population. The recommended levels for the KALPHA to be considered a good inter-rater reliability is above 80% and a poor inter-rater reliability is below 65% (De Swert, 2012).

The use of the statistical data analysis KALPHA in SPSS is appropriate because it calculates the inter-rater/inter-coder reliability for coefficient for multiple coders using at least nominal/categorical level data (Freelon, 2010). The BRT categories being used to rate the student captures are categories in a distinct order, thus they are ordinal data. Using this analysis allows for the analysis of multiple variables in this study, two or more teacher evaluations of the same student's work. The participants' scores were also compared to all the other participants to check for the statistical probability that any scores were due to chance. This analysis used a categorical/nominal variable for each participant's name. An ordinal variable represented each student score provided by the teacher participants. Although these BRT scores did not occur at exact intervals, they did occur in an ordinal manner. This ordinal analysis utility is based on the Kappa Coefficient, which pairs all the coded student samples with the teacher raters to the scores assigned (Krippendorff, 2011). Kippendorff's Kappa Coefficient formulas permit the analysis of more than one piece of work per student. This statistical analysis approach best fit with this study's multiple student, multiple writing captures.

Table 7

Krippendorff's Alpha Reliability Estimate

	Alpha	LL95%CI	UL95%CI	Units	Observers	Pairs
Ordinal						

Notes. Abbreviations: Lower Level Confidence Interval (LL % CI), Upper Level Confidence Interval (UL % CI)

Assumptions, Limitations, Scope, and Delimitations

Assumptions

The capture and scoring of student writing samples yielded data that could be used to enhance teachers' understanding of student thinking. Using these data teachers could conceivably be able to identify what levels of thinking their students are proficient and would be able to bridge gaps between all students using formative assessments to impact future instruction depending on what students needed to develop. Additionally, I assumed that student writing was given a rating by the teachers that reflected their best effort.

Limitations

The lack of a larger teacher participant pool is a limitation for this study. While the study includes an acceptable number of participants, the results of the inferential

statistical analysis provide only internal validity. When a non-random sample is representative (when characteristics of the sample are comparable with the target population) the results are generalizable (Banerjee & Chaudhury, 2010). While data from purposive, non-random samples is likely not generalizable to larger populations, it may provide hints for future random studies that would be generalizable to the larger population. Asking the participants to code additional student samples could increase the validity of the study. However, it is unlikely that participants would want to volunteer for longer than the approximately 5 hours the study will take to score the 52 writing samples.

The use of technology as part of the collection process for raters' scores poses further potential limitations. While unlikely, technical issues may arise during the use of Google sheets to collect the rater scores based on the requirement for internet connectivity while scores are recorded. Issues with internet connectivity is not expected due to the widespread use and availability of internet.

Scope and Delimitations

In this study I investigated the inter-rater reliability of using BRT as a rubric to score samples of students' writing stored in a digital portfolio software. The writing will be scored to identify varied levels of thinking in each writing sample. The study was delimited to 52 samples of seventh grade writing that was be scored by 8 teachers. This study encompassed the writing of seventh grade students. This study included middle school teachers in all content areas within two educational organizations ranging through the fifth, sixth, seventh, and eighth grades.

Protection of Participants' Rights

This study relied on middle school teacher participants and archival student work that the teachers evaluated. All teacher participants of this study will be voluntary. As the principal researcher, I will host a meeting at each school in which she will communicate the purpose and process of the study. During this meeting, and after it via email, volunteers were able to sign up for participation in this study. Participants signing up and voluntarily attending the proposed session received a $20 stipend Starbucks gift card paid by myself. Additionally, the study participants were provided with three breaks. Participants were provided two snacks and one lunch during these breaks. Drinks were readily available during the training and coding sessions. Restrooms were readily available throughout the entire training and coding process.

Participants signed up to participate in the one hour training session, in addition to the approximately four-hour coding session. Participants arrived and were greeted with a beverage of water, coffee, and or tea. The training session lasted for 1 hour. After 1 hour participants were provided with a snack break for 20 minutes. The scoring session began after participants returned to the designated area. Once participants returned the coding process began. After each hour spent scoring, participants were provided with a snack break of 20 minutes. Scoring the 52 captures took approximately four hours of time. After the scoring was complete, participants were provided with lunch. At this time, all participants were given a stipend gift card. The day took approximately five to six hours total.

Participant protections were ensured through the granting of permission for this research study by Walden University's Institutional Review Board (IRB) prior to collecting any data. There is no record of participant names and all data were coded to ensure the de-identification of participants and students. Identification numbers delineate all samples of student writing. All data collection has been kept confidential as the participants and researcher spent the day together, but their data was entered for analysis using 5-digit numerical identification codes. All participants signed a consent form prior to participation in the project study.

Although I work in the larger educational charter school system from which this data collection occurred, the middle school teacher participants work in a different building and are supervised and evaluated entirely by another administrator. Additionally, all measures were be taken to keep the data collected entirely confidential and only linked through assigned 5-digit identification number. Participants scoring the confidential writing samples were not affected in any way by the scoring process as there is no link between the establishment of inter-rater reliability and teacher or student performance.

Data Analysis Results

The research for this project study was conducted through a scoring process in which participants completed a brief training for scoring using the BRT. After the brief training and warm up exercises, the participants scored 52 writing samples, two samples taken from 26 seventh grade students. The data were recorded using Google Sheets, each of which were associated with a confidential participant number. Once all data were entered into the Google Sheet associated with each of the confidential participant codes I

was able to access the data to begin analysis. An additional spreadsheet was created which included the pre-existing grade given to each of the writing samples. These grades had been removed from the writing sample prior to the scoring session and the writing samples were also de-identified.

Descriptive Statistics

Once the data had been collected, I transferred the data into the SPSS spreadsheet to prepare for analysis. The data analysis began with descriptive statistics to identify the mean and then mode ratings for the 52 writing pieces scored by each participant rater for levels of thinking using the BRT, which are presented in Table 5. The data were analyzed to demonstrate how many writing samples were scored if they included writing in which HOTs were evident, if the scores demonstrated that the writing contained mostly LOTs or scored to demonstrate that the writing included both a mix of HOTs and LOTs. The participants mean, mode, and standard deviations for levels of thinking in ascending order are shared in table 8.

Table 8

Participant Mean, Mode, Scores for Levels of Thinking in Ascending Order

	Participant	Mean	Mode
1	PART4	3.62	2
2	PART6	3.75	3
3	PART3	3.77	3
4	PART7	3.88	3
5	PART1	4	4
6	PART8	4.5	6
7	PART2	5.12	6
8	PART5	5.25	5
Total		8	8

Research Question 1

What is the relationship between teacher grades for each writing sample and the BRT-based LOTS - HOTS ratings for each writing sample?

This research question was aimed at determining the discriminant validity of the scoring rubric. If the grades teachers assigned to the writing sample were not statistically different than the ratings participants assigned to the same writing sample, then this would be evidence that the original teacher grades and the rubric evaluations were evaluating essentially different constructs. The validation of the BRT scoring rubric was

a basic premise for this study. The categorical BRT ratings were compared to the existing teacher grades in SPSS using a Spearman rank correlation. According to Meghanathan and He (2016), correlation ranges are .00 to .19 are very weak positive, .20 to .39 are weak positive, .40 to .59 are a moderate positive, .60 to .79 are a strong positive, and .80 to 1.00 are a very strong positive.

The data from this study demonstrated no relationship between the classroom grades and the BRT ratings assigned to each writing sample. A Spearman's rank-order correlation was run to assess the relationship between using the BRT as a scoring tool to determine levels of thinking evident in student writing samples and the existing teacher grades that had been assigned preceding the study. There was no statistically significant correlation between the existing teacher ratings and the scores from the BRT rubric, r_s (49) = .365, p =0.01. Indeed, this was the outcome that would have served as evidence that the teacher grades and BRT rubric ratings were measuring two distinct entities, thus providing evidence of discriminant validity for the BRT rubric. This finding does not serve as evidence of discriminant validity, however, because the BRT ratings were statistically unreliable as the next sections will explain. After the BRT was deemed reliable, it could be a worthwhile endeavor to re-examine the relationship between pre-existing grades and scores using the BRT.

Research Question 2

Will there be moderate (> .7) or higher inter-rater reliability demonstrated by middle school teachers' ratings using the BRT rubric for scoring writing samples of student's demonstrations of thinking? I examined the data for inter-rater reliability using

inferential statistics to determine the BRT's reliability using the Krippendorff's Inter-rater Reliability Estimate, which uses a point estimate of the inter-rater reliability. In order to analyze the data set, I added in the following macro syntax: Kalpha judges = V1 V2 V3 V4 V5 V6 V7 V8 / level2 / detail = 0 / boot = 10000 within the SPSS software. This macro instructed SPSS to use the Krippendorff's Kappa Coefficient formula to analyze multiple writing samples. This data analysis revealed that the use of the BRT to score writing for levels of student thinking was not reliable (see Table 9).

Table 9

Krippendorf's Alpha Reliability Estimate

	Alpha	LL95%CI	UL95%CI	Units	Observers	Pairs
Ordinal	.0533	-.0245	.1308	52.000	8.000	1456.000

Notes. Abbreviations: Lower Level Confidence Interval (LL % CI), Upper Level Confidence Interval (UL % CI)

The data supported the second null hypothesis that there will not be a moderate (>.7) or better inter-rater reliability based on Krippendorff estimates of middle school teachers' ratings using the BRT rubric for scoring multiple writing samples of levels of student thinking. The findings from this study revealed that there was not a moderate (>.7) or better inter-rater reliability demonstrated by the middle school teachers' ratings using the BRT rubric for scoring thinking levels within the student writing samples.

Research Question 3

Will there be a moderate (>.7) or higher inter-rater reliability demonstrated by the true-population Krippendorff alpha estimates between middle school teachers' ratings using the BRT scoring rubric for multiple samples of student's levels of thinking?

The data supported H3o. There was no moderate (>.7) or higher inter-rater reliability demonstrated by the true population Krippendorff alpha estimates between middle school teachers' ratings using the BRT scoring rubric for multiple samples of students' levels of thinking.

Summary of Findings

Unfortunately, the BRT as a scoring rubric was not reliable based on this examination. There are a number of variables that could have impacted the lack of reliability of the BRT as a scoring rubric. For example, in order for the BRT based rubric a to be reevaluated for reliability for use as a scoring tool, it would need to be improved upon. For example, the content of the BRT could be rearranged into a smaller number of descriptors for HOTs success.

While it did not make sense to evaluate the discriminant validity of a rubric that did not reliably assess levels of student thinking, the analysis was completed to fulfill the obligations of the project study. There was no statistically significant relationship between using the BRT as a scoring tool to determine levels of thinking evident in student writing samples and the existing teacher grades that had been previously assigned. The originally scored writing samples were scored based on a rubric that focused on published writing. The inter-rater reliability of the original rubric is unknown,

and may have been similarly poor, a situation that may have contributed to the lack of correlation found. At any rate, this investigation of the correlation between the original grades and the BRT scores assigned to assess student thinking were not related.

The BRT as a rubric to evaluate student thinking could have been flawed in its structure and scoring as it has not been previously determined reliable or valid for scoring levels of thinking. The BRT is largely used to plan for instructional tasks in which student potentially reached higher level of cognition as associated with intentional learning experiences (Steedle & Ferrara, 2016). For example, the Peak to Peak Center for Professional Development trains educators to utilize a condensed version of the BRT as recommended by The College Board (personal communication, Director of Professional Development, 2018). While this program only trains teachers to use the BRT to plan instructional tasks, it is possible that their version would be better to use as a rubric than the one used in this study. This version divides levels of thinking into three categories of cognition including (a) Level 1 – factual information that can be looked up in a book, (b) Level 2 – the why or the how which takes the thinking to a procedural level of understanding, applying, and analyzing, and (c) Level 3 – the universal (human connection) level of conditional knowledge including the *why does this matter* levels that include evaluating and creating (personal communication, Director of Professional Development, 2018). Designing a rubric for thinking in a more finite manner may contribute to the development of a more straightforward rubric based on the BRT that is both reliable and valid.

In order to contribute to designing an objective framework other than BRT, it is important to consider alternative options for the assessment of thinking. One possibility could be Epstein's (1998) cognitive-experiential self-theory (CEST), which currently includes a valid and reliable measure of thinking, and could contribute to designing an objective framework from which to build new rubrics to grade thinking. Using the valid construct of the CEST, researchers could propose the heuristic process of constructive thinking as a framework from which to base the rubric design of a scoring tool for sound thinking (Epstein & Meier, 1989). Healthy, constructive thinking, which includes the absence of mal-adaptive thinking (Epstein, 1998); however, is not necessarily higher level thinking as conceptualized in Bloom's taxonomy.

While it is possible that the investment into more comprehensive training of participants could result in an increase in the reliability and validity of the BRT as a rubric, such an investment would be ill-advised without research-derived rationale for pursuing that solution. It is possible that the BRT as a rubric would need to be improved upon prior to increasing its reliability and validity as a scoring tool for levels of HOTs.

Project Deliverables Based on Findings

With the approval of my committee the project deliverable included a white paper discussing the research study, its shortcomings, and potential pursuits for further research design. To meet the requirements of a position paper, I selected a white paper to complete this project. The intent of the white paper is to inform interested stakeholders within my learning community about the findings of this research. Additionally, a goal of the project is to explore other avenues for accurately assessing levels of thinking that include the

voluntary contributions from stakeholders on how to redesign the BRT as a rubric to make it more reliable and valid.

Conclusion

The goal of this study was to fill a gap in educator practice for scoring student writing to include levels of student thinking using the BRT rubric. There is a risk when proposing a study that it is not grounded in familiarity or common practice, such as evaluating student thinking levels. The risk is that the study could be rejected by the participants. This risk was mitigated in two important ways. First, the development of higher order thinking skills is an espoused value of the involved schools. Second, the value is also an educational goal that is highly supported by the teachers who work at the schools. Teachers, however, would like to know that such evaluations are reliable and valid, and that desire reflects the purpose of this study. Teachers want to be able to evaluate how well their students are thinking. Teachers want to prepare students for success in the 21st century global economy. Based on this gap in practice, this research design will provide feedback on whether or not the BRT rubric is valid and reliable. Care has been taken to ensure that all ethical considerations have been addressed and planned for.

Section 3: The Project

Introduction

Section 3 includes information about the project study. In this section, I provide a rationale for my project study selection, a project description, an evaluation of the project, and a discussion of the implications of the project at its culmination. This project is the delivery of a white (i.e., position) paper written for stakeholders within my learning community. The purpose of the white paper is to share the findings from the project study in an applied format that is more consumable for the practitioner.

The white paper provided in Appendix A includes background about the existing problem within the field of education of the lack of reliable and valid assessment tools to evaluate students for HOTs. The purpose of the paper was to provide a brief of the study findings and recommendations for consumption by education practitioners. Based on the additional review of literature, I included information in the white paper on seven important areas for evaluating higher-order thinking, including (a) BRT, (b) types of thinking, (c) learning environments, (d) 21st century learning, (e) HOTs and LOTs, (f) assessment, and (g) rubrics. Finally, in the white paper I outline assumptions and offer recommendations based on my research study results and the research literature--for example, suggested revisions to the BRT to make it a more reliable and valid rubric for scoring student writing for HOTs and LOTs.

Rationale

Using the findings from the study, I developed a position paper to convey my assumptions and recommendations for future research. I had originally considered using

two different approaches for presenting this information, but, based on the research study results, I concluded that neither professional development nor curriculum development training would have been appropriate. In the event that the data demonstrated that the BRT was a reliable and valid tool for scoring student writing for levels of thinking, both professional development or curriculum training would have been appropriate. Because the findings did not show that the BRT rubric was a reliable tool for scoring student thinking, I concluded that it was unwise to provide professional development on its use at the present time. Instead, I determined that the white paper would be the most appropriate way to share the research on the current state of the literature and the difficulties I encountered when evaluating BRT as a potentially valid and reliable tool.

The white paper was a medium through which I provided stakeholders at my study site and within my learning community with research-based information on scoring writing for evidence of varied student thinking levels. I also wrote the white paper to inform and possibly prompt further research to continue exploring a reliable and valid method for assessing student writing for varied levels of thinking. The assessment of students' levels of thinking could be used to promote further instruction to ensure students develop these necessary skills before graduation.

Review of the Literature

When gathering review for this literature review I focused on search terms which would provide insight on potential project directions. I searched peer-reviewed articles, journals, and dissertations using the Walden University Library and including the following databases: Sage Journals, Taylor and Francis, and ERIC. Some search terms

used were *professional development*, *white papers*, *policy recommendations*, *HOTs and LOTs*, *rubrics*, and *assessments.* After searching the following themes emerged*:* program evaluation and document analysis, white papers as a method of prompting future action in the field, professional development, and scoring writing.

Policy Recommendations

During this literature review the most prominent theme to surface focused on the use of research writing to prompt future action in the field. Hassel et al. (2015) identified the use of white papers within a field as a method of presenting current research and making recommendations to professionals in the field. The TYCA authors of the white paper used data collected from a case study on writing courses at 2-year colleges to illuminate placement practice (Hassel et al., 2015). This white paper related to my study because of its focus on social change within the field of education based on the proposal for best practices. A theme in much of the current literature within the field of education is that there should be a reevaluation of the purpose for education and thus a rethinking of the best pathways to achieve necessary reforms through research-based policy recommendations (Hassel et al., 2015).

Roberts-Mahoney, Means, and Garrison (2016) noted the use of policy recommendation in a white paper after analyzing content on personalized learning technology. Roberts-Mahoney et al. study employed purposive (or relevance) sampling and Krippendorff's alpha coefficient as a statistical measure. The aim of the study was to frame the purpose of public education based on recent document analysis within the field. The researchers initially sampled documents in various formats although each of the 12

documents was considered one unit. The researchers used four thematic questions to code and analyze the data from each unit. Roberts-Mahoney et al. ended the white paper with a prompt for a comprehensive rethinking of the purpose of education through the evaluation of potential best practices and innovations within the field. Similar to my project study, this white paper acts as a call to action around the need for evolution in the practice of designing and use of assessments focused on evidencing deeper levels of learning.

Sotiriou, Riviou, Cherouvis, Chelioti, and Bogner (2016) examined the introduction of large-scale innovation through a white paper discussing the program evaluation of tech supported innovation through a three-phase innovation scheme. The study included participants from 400 schools and yielded four statistically significant themes, with a final evaluation that the school innovation model yielded apparent positive results (Sotiriou et al., 2016). This type of innovation supports current literature regarding the need for the implementation of research based assessment tools to which would demonstrate students capacity for HOTS.

Candal's and Pioneer Institute for Public Policy Research (2015) white paper discussed case studies from five high performing charter schools in Massachusetts and recommended transitioning the focus from highly qualified teachers to teacher effectiveness. The study recommendations included the following themes: teacher effectiveness, the important of hiring, promoting excellence through modeling and feedback, and the evaluation of student performance (Candal and Pioneer Institute for Public Policy Research, 2015). The authors of the Education Excellence Everywhere

White Paper (2016) additionally discussed the evolution of necessary policies and structures in place to ensure the maintenance of highly effective teachers. The focus of this white paper on the evaluation of student performance lends itself to the pursuit of assessment tools designed to discern students' capacity for thinking at deeper levels.

Jimerson and Childs (2017) noted the influences on educational policy in a white paper. This white paper recommendation focused on the use of data trends as signals which should determine what actions need to be taken and commitments made by policy makers to obtain the ideal outcomes symbolized within the field of education (Jimerson & Childs, 2017). Educational data use informs policy actors, who must use the signals of effective data trends to frame expectations that align with research to impact practice in an effective way (Jimerson & Childs, 2017). To make necessary changes in educational policy more research must be conducted to shine a light on the need for reform of assessment tools which can better address students' capacity for 21st century skills such as the application of HOTS.

Within the field of education, there are a number of different white paper formats (Campbell & Naidoo, 2017). Cullen (2018) identified a white paper as an authoritative document used to inform the reader with expert knowledge or research to propose a solution or recommendation. Other purposes for white papers include conveying policy, presenting tech information, sharing information on a completed project to propose future projects, or, in recent years, sharing information for marketing purposes (Hyde, Stolley, & Sakamuno, 2015). The white paper written using the results of my study is rooted in

the dissemination of a call to action regarding the continued evaluation and implementation of rubrics to assess levels of student thinking.

HOTs and LOTs

The development of HOTs is essential for students to reach their potential to become effective contributing members of society as adults. Developing the capacity to solve everyday problems and establish solutions when faced with a challenge is not something that is currently taught in traditional school systems in the United States (Wiliam, 2011). Traditional schooling models primarily utilize the bottom levels of the BRT and fail to bridge the gap between the concepts and content learned and the HOTs necessary to use them (Kaldor, 2018). Scott (2017) delineated three main frameworks of 21st-century skills, including (a) learning and innovation skills, (b) life and career skills, and (c) information, media, and technology skills of which HOTs are grouped under the learning and innovation skills. Additionally, Ganapathy and Wai Kit (2017) supported that the focus of traditional school systems is the reproduction of knowledge versus the manipulation transformation of information that occurs when a student is working in the three upper levels of cognitive skills in the BRT: analysis, synthesis, and evaluation. In order for there to be necessary change in the national vision regarding these deficits, policymakers must acknowledge the failures of the current system and must make adjustments that match the evolved expectations for students in the 21st century competing for employment.

Recent research on the development and assessment of HOTs proposes the engagement of students in their learning in active learning and student-centered ways

(Jones, 2012). Retnawati, Djidu, Kartianomi, Apino and Anazifa (2018) discussed the element of synergy between stakeholders in the pursuit of training teachers to train students in HOTs. To effectively implement the development of HOTs, teachers, curricular updates, and the continued development of teaching professionals must be at the forefront of this critical implementation (Purnomo, 2017). Problem based learning, discovery learning, inquiry based learning, and any model using contextualized problems will provide the necessary training experiences for students through which they can develop HOTs (Retnawati et al., 2018).

Bartell (2013) proposed that teachers can achieve these types of experiences within their practice by playing an active role in planning, implementing and evaluating HOTs oriented learning. A challenge in the implementation of HOTs-based learning experiences is the misunderstandings that teacher generally have around the types of learning opportunities that could be used to train students for HOTs. While teachers generally value HOTs as the skills students need to solve everyday problems, they are unable to articulate the steps of operational implementation of the necessary learning experiences (Jailani & Retnawati, 2016). According to Jailani & Retnawati (2016) teachers have identified methods for the assessment of HOTs such as contextual based essay prompts, but have not found the link between the measurement of HOTs using the BRT in which they note HOTs as the top three categories: analysis, synthesis, and evaluation. To move forward, educators must develop a clear understanding of HOTs and how to develop, implement, and assess HOTs in order to train students adequately.

Assessment

The absence of a generalizable framework or assessment tool that measures student thinking through writing is the deficit within the field of education. On a large scale, assessment design does not indicate 21st century learning goals including thinking skills (Brown, 2016). Students must demonstrate competencies in critical thinking, problem solving, collaboration, and autonomous independent transfer of knowledge to exercise HOTs (Wagner, 2014). Assessment data focused on thinking and learning must regularly be collected, must inform instruction, and must be pulled from a pool of success criteria universal to the learning community (Brookhart & Chen, 2015; Moss & Brookhart, 2009).

Epstein and Meier (1989) published the Cognitive-Experiential Self Theory (CEST) to measure thinking patterns underlying emotional wellbeing. The CEST examines three independent thought systems: the rational system, the experiential system, and the associationistic system (Epstein & Meier, 1989). This theory of personality aimed at the understanding of practical intelligence assumes that everyday perception and behaviors are influenced and organized mainly by the experiential conceptual system (Epstein & Meier, 1989). The Constructive Thinking Inventory (CTI) was designed out of the desire to understand the experiential system as a measure of intelligence as it was the key system in regulating practical intelligence (Epstein & Meier, 1989). While these measures give insight into the emotional wellbeing and practical intelligence of student's behaviors, the CTI does not provide an evaluation of HOTs and LOTs within student writing.

Fortunately, based on the deficits in the HOTs and LOTs that businesses and colleges have identified the assessment of thinking has begun to surface in state standardized assessments. Such standardized assessment tools include Partnership for Assessment of Readiness for College and Career (PARCC) and Smarter Balanced Assessments in the form of Performance-Based Assessment components (Benjamin et al., 2012; Herman et al., 2013). Considering that large-scale testing corporations' attention has begun to focus on this area of need, policy makers and stakeholders informing practice in the field have started to take notice.

This subjectivity inherent in assessment of written work manifests through practice in many ways. Hess et al. (2009) noted the discrepancies in teacher scoring as they fall into old habits of scoring on academic enablers such as student past behaviors or achievements. For example, if a teacher scores an essay and the rubric is vague they are likely to factor in historical subjective observations and associations from interactions with the student in the past. Additionally, a student's actual academic competence and habits may factor into what a teacher identifies or focuses on if, for example, the student is frequently late in turning in work. The attitude of the teacher may be less open to possibilities of the range of LOTs to HOTs in each student's writing.

While a good deal of research is available related to classroom assessment, there remains a gap in research around the documentation of student LOTs and HOTs (Wiggins & McTighe, 1998). McMillan (2013) identified the need to develop principles of assessment that document student learning, addressing specifically the necessity for a supporting body of research on classroom assessment. Furthermore, discussed in research

is the need for developing evidence of in-depth descriptions of how teachers summarize and document learning and how that learning progresses (Brookhart, 2013; Vanlommel & Schildkamp, 2018). In the 21st century, with the transformation in the contexts for assessment, Aagaard and Lund (2013) identified the lack of educator's experience in how to assess collaborative and interactively constructed learning (p. 223).

There is a difference between design of learning questions and the assessment of the thinking generated from those questions. While there are structures in place, such as BRT, to guide the design of learning opportunities and questions that address higher order thinking, there is a breakdown in the assessment of the responses to the questions focused on higher order thinking (Vista et al., 2015). Bøhn (2018) discussed his research in which teachers are familiar with the assessment of the what (knowledge) but are unfamiliar with the how (cognition) which calls for the further development of teachers to understand this difference and begin to develop assessment tools that evaluate student's abilities to present their discoveries.

In the design of evaluation tools to use in a formative or summative manner, indicators must be identified for ideal student outcomes for specific tasks (ideally HOTs based opportunities) within a given discipline. For example, Atherton (2013) discussed the phases of learning using a Structure of Observed Learning Outcomes taxonomy, in which indicators are checked off as the students' learning progresses through Piagetian developmental phases beginning with the pre-structural through the extended abstract level in which students transfer from simple to complex applications. Raiyn and Tilchin (2016) proposed a method for the adaptive complex assessment of HOTs through a

problem base learning process. This is a three-stage assessment process that prompts the development of HOTs through each stage of (a) developing the HOTs, (b) developing the HOTs and collaborative skills, and (C)) assessment of the collaborative skills and construction of summative assessments of students (Raiyn & Tilchin, 2016). In addition to the PBL process which is student centered and adaptive in ways that allow for the development of students' HOTs, researchers have also discussed the necessity of student engagement through choice and flexible assessments as methods through which students develop necessary HOTs.

Pretorius, van Mourik and Barratt (2017) proposed the development of flexible, student choice based assessment through which students are offered options and choose which to pursue. Biggs (2012) noted that student engagement and buy in are considered central to effective educational practice. Authentic assessment task options presented to students allow them to see the transferability of skills being assessed to their future applications (Pretorius et al., 2017). When Pretorius et al. (2017) evaluated assessments based on both product-focused activities and process-focused activities, the assessment tools from the process focused (PBL type activities) were more effective in prompting deeper levels of (HOTs) thinking.

Through the careful examination of best practice in assessments and feedback regularly provided to students, educators can begin to address the gaps in practice of the assessment of thinking. It is no longer an option to assess students using an unbalanced approach in which only LOTS are assessed using traditional standardized and summative measures. Educators must design learning opportunities that demonstrate students

thinking capacity and their ability to apply what they have learning in a variety of setting and for a variety of purposes. These types of reflexive assessments and rubrics for the assessment of thinking can propel students to competencies in skills needed for the 21st Century and competition in a global economy (Dawson, 2015).

Rubrics

The research on training teachers to use rubrics clearly demonstrates the need for comprehensive training in the use of rubrics to ensure the positive effects of rater reliability. While the study results from this research did not deem the BRT reliable or valid, it is possible that with some improvements, it may be reevaluated and found reliable and valid. Taylor and Galaczi (2011) discuss the need for comprehensive teacher training in rubrics based on the element of perception and the need to clarify evidence in student work when compared with rubric criteria. Often questioned in current research is how well teachers understand the constructs that are being assessed using a rubric and how this is an additional area in which teacher training is required when assessing students using criteria-based rubrics (Yildiz, 2011). Bøhn (2018) maintained that teachers as raters using rubrics effectively, can significantly impact student learning opportunities to establish genuine learning around HOTs.

The research is also clear on the importance of using rubrics. When comparing the benefits of rubrics to comprehensive graded category rating scales, Dogan and Uluman (2017) found that rubrics provide better access to consistent, genuine, formative assessment as a method of student feedback. Hassel (2015) found that measuring student learning in a manner that provides clear criteria (a rubric) makes

visible the measurement of student thinking and learning. While this is a more time-consuming method of measurement and is not without challenge, the tradeoff of effort is worthwhile as once the levels of expected proficiency have been delineated, genuine assessment of learning and progress can occur in an objective fashion.

The research is clear that rubrics can and should be used to measure HOTs. For example, Rembach and Dison (2016) studied the transformation of taxonomies into rubrics and demonstrated learning benefits in determining student's cognitive capacity when faced with set tasks. Constructive alignments (CA) between course descriptions, learning objectives, teaching and learning, and assessment must be interrelated for deep learning to occur (Biggs & Tang, 2011). Furthermore, Rembach and Dison (2016) note the promotion of HOTs when teachers, scorers, and students, had access to rubrics all the time to use as a feedback tool to gauge progress. Using rubrics to determine levels of student thinking in combination with learning structures designed for authentic learning is imperative to the successful evaluation of student's competencies (Hohmann & Grillo, 2014). This type of interconnected planning and assessment requires that educators are trained in a comprehensive manner with opportunities for coaching and mentoring through continue professional development and collaborative efforts.

Professional Development

In an effort to determine methods for implementing next steps in the field based on the evaluation of current research, I examined literature on professional development design and best practice. Jacobson (2016) emphasized the importance of scaffolded sessions, which are presented in a variety of structures that support discourse among

collaborating educators. While Derrington and Kirk (2017) focused a case study on the efficacy of job-embedded professional development by collecting data from interviews of participants at 28 K-12 schools. After the completion of the coding process a master list emerged regarding effective job embedded professional development highlighting a call for professional development to be learner centered, knowledge centered, community centered, and assessment centered. Lauer, Christopher, Firpo-Triplett, and Buchting (2014) reviewed literature which echoed the necessity for professional development to be focused on participant outcomes through the focus on professional development design being learner centered.

Project Description

The project for this research study was a position paper that was shared with stakeholders in my learning community and local community. The findings shared in this position paper are a stimulus for continued study of the assessment of HOTs and LOTs within my learning community. It is the goal of the white paper to act as a catalyst for the continued pursuit of best practices in preparing our students for 21st-century competition in a global and local society.

The white paper shared with stakeholders of the findings of this research, of the continued need for evaluation of student thinking, the possibility of using a rubric based on Blooms Revised Taxonomy (BRT) to score levels of thinking in writing samples, and the need for substantial professional development of teachers to utilize a BRT based rubric if it is found to be a reliable and valid tool in future studies. My further research recommendations in the white paper focus on the need for professional development on

utilizing a BRT as a rubric in the event that once is designed and found to be a reliable and valid scoring tool. This is consistent with other research on using rubrics to score student written work (Holt et al., 2015).

The white paper as a project is intended to be emailed to stakeholders and those within the learning community, therefore, limited resources will be needed for the project. I will need a computer, access to the internet, as well as the email addresses for the institutions and community directories to which the project will be emailed. Potential barriers to the dissemination of this project will be the accuracy of emails recorded in directories of stakeholders and those within the learning community. Having access to the newsletter posting for both my charter school directory and the other charter school directory will provide a solution to this potential barrier. This project will be emailed once final project acceptance is received from Walden University's Chief Academic Officer. Upon emailing the project, the accompanying evaluation link will begin to auto generate based on the readership of the white paper and feedback stakeholders provide. My role in this project will be to disseminate the white paper to stakeholders from both charter schools and within my learning community.

The white paper will be emailed to key stakeholders within my local learning community such as school board members, the council for our municipality, our parent body, teachers, school leaders and additional coalitions and outreach programs within my local community. This white paper will also be emailed to the faculty (via the director of professional development) of the public charter school with which we share educational practice around development of students thinking capacity. This charter school is a

regional professional development training center for charter schools in the state of Colorado and is in perpetual pursuit of best practices and remains interested in how current research impacts the field.

Once the white paper has been emailed to the key stakeholders I will analyze the project evaluation feedback to guide next steps in the continued exploration of the BRT as a reliable and valid rubric for scoring thinking in writing samples. Using the information collected from the project evaluation as well as the results from this study, I will continue to analyze ways to improve the BRT as a rubric for scoring writing. For example, categorizing the BRT levels into three groups encompassing the evidence from the varied levels included. An additional adjustment in addition to improving upon the BRT could be the enhancement of job-embedded professional development opportunities during which educators norming the process for scoring writing using the BRT.

Project Evaluation Plan

To evaluate this project, I will share a Google Form questionnaire with all stakeholders with whom the position paper was shared. The voluntary one-item questionnaire requests that stakeholders provide suggestions that would help make the BRT rubric more valid and reliable. The results from this questionnaire automatically pool into a Google Sheet linked to the Google Forms questionnaire. The results of the questionnaire provided stakeholder input on further pursuit of the BRT-based rubric for scoring varied levels of thinking through student writing samples. Collecting suggestions from stakeholders regarding methods of improving the BRT for scoring will likely elicit a range of suggestions through which those focused on best practice and current research

will be most valuable. Key stakeholders whose project evaluations would be most useful will be educators, school leaders, or those who inform policy within the learning community.

Project Implications

This project provided a starting point for the continued development of teacher's awareness of HOTs and LOTs as well as the continued professional development of teacher's capacity for providing learning experiences in which students can develop these HOTs and evaluate student success. Furthermore, this project aimed to build awareness and interest in the field around the use of a rubric to score student levels of cognition within writing. Additionally, this study has provided a starting point of data which could be used to modify and improve the rubric from which point another validation study could be conducted to see if the modified version is any more valid and reliable than the first.

While the implications of the study are largely a body of evidence positioned as a starting point for the continued redesign of the BRT as a rubric for scoring thinking, the factors preceding reevaluation in further study of this, may include a more comprehensive preparation program. Building teacher awareness of the BRT and students varied levels of thought to ensure a firm understanding of the importance of HOTs and LOTs is a critical touchpoint before educators are able to articulate the scaffolded implementations necessary for students to develop these skills. Once an educator is able to make this articulation, the focus should shift to the development of a BRT based rubric

as a scoring tool and the sustained and evaluated professional development opportunities for teachers to practice implementation and use of the rubric to score writing for thinking.

Section 4: Reflections and Conclusions

Introduction

In this section, I discuss my study on the use of the BRT as a rubric for assessing seventh-grade student thinking. The project study purpose was to determine if the BRT, as a rubric, would be a reliable and valid scoring tool to evaluate student writing samples for varied levels of thinking. Participants in the study included eight middle school teachers from public charter schools, both of which focus on developing HOTs. Participants engaged in a brief training in the use of BRT as a scoring rubric for student thinking. The findings from the research led me to develop a white paper to distribute to local stakeholders in my learning community as well as the other charter school from which participants were invited.

I used a quantitative, nonexperimental research design to investigate the discriminant validity of the BRT as a rubric and the interrater reliability between teachers scoring student samples of writing. This methodology allowed me to evaluate the variables in the context in which they naturally occur. Phelan and Wren (2006) hold that interrater reliability assesses the degree to which two raters agree in their determination of a score. Therefore, I quantitatively analyzed teachers' use of the BRT as assessment rubric criteria to evaluate students' levels of thinking using archived writing samples. Participants scored writing samples using confidential Google Sheet logins to input their scores based on a scale ranging from one to six (1) *Remembering*, (2) *Understanding*, (3) *Applying*, (4) *Analyzing*, (5) *Evaluating*, and (6) *Creating* associated with the six BRT levels.

In analyzing data, I determined that the BRT is not reliable or valid as a rubric in the circumstances of my project study which offered limited teacher training based on the BRT rubric. The information gathered from my data collection provided a very clear direction regarding necessary components in professional development around teachers' capacity to utilize the BRT as a rubric for assessing thinking. The white paper includes background information about the study, in addition to a discussion of the challenges associated with educator awareness and implementation of the BRT both to design learning opportunities in which student can develop HOTs, as well as using BRT as rubric criteria with which to evaluate thinking.

In this section, I reflect on the BRT as a reliable and valid rubric for scoring student thinking. I address how a BRT-based scoring tool might be a benefit to those designing, implementing, and assessing learning opportunities in which students develop and demonstrate HOTs. I also speak to the strengths and limitations of my project study, offer recommendations of future research, and deliberate the propositions of my research.

Project Strengths and Limitations

Strengths

The strength of my project stems from the drafting of a position paper that requests further evaluation of examples of educational practice of the competencies listed in the Organization for Economic Co-operation and Development (OCED) Learning Framework 2030 (OCED, 2018). OCED (2018) identified five challenges commonly found within the field of education, noting the impact that the level of content has on a student's ability to authentically engage in the learning process and to reach deeper levels

of learning. My project deliverable, which is in the form of a position paper on the necessity for the study of the assessment of HOTs and LOTs study, is rooted in the constructivist framework and focuses on the examination of the BRT as a potential tool that could be used to reliably and validly assess students' thinking capacity in a measurable manner.

Additionally, the position paper I wrote reinforces the necessity of continuing to explore alternatives for assessment ultimately focused on students' HOTs capacity. Soland, Hamilton, and Stecher (2013) discussed the use of interim assessment that provides actionable information based on a student's demonstration of skills such as critical thinking. There is a lack of research focused on the lack of assessment and feedback regarding the stages of development for 21st century cognitive competencies (Soland et al., 2013). The authors of the OCED Education 2030 project asked for a reorientation of the purpose and intention of education and specifically discussed the need for contributions from researchers and experts to strengthen this need for change (OCED, 2018). My project study strengthens this call to action for the continued investigation of reliable and valid assessment tools for the assessment of HOTs and LOTs.

Limitations

Although I did not determine that the BRT was a reliable and valid tool for scoring writing for varied thinking levels, I was able to provide data in the project white paper that may spare another researcher spending time pursuing the same research. Another researcher may find the references in my white paper to be a viable starting point

for continued research. The white paper contributes to the field through the recommendation of next steps in the evaluation of the BRT as a reliable and valid rubric for scoring thinking.

I have identified three limitations of my project. The first is the likelihood that within my small learning community that not many people will be pursuing postgraduate research that would be published on a more global scale to be later accessible in the field. Teachers, administrators, and parents of students within my learning community may not be willing to begin research that directly picks up where my study left off and where the white paper makes recommendations for future research. Additionally, those stakeholders not directly involved in the development and assessment of HOTs may be more focused on the remedial pressures of the learning community. For example, they may be more likely to pursue the enhancement of students' achievement scores if they are below grade level.

The second limitation of my project is a significant lack of funding within my learning community. As an independent, public charter school, not governed by the local school district and therefore not eligible for receiving the same funding that the local school district receives in the way of the local tax monies. Based on this funding disparity in per pupil revenue, my learning community is likely unable to bridge the gap in funding and therefore is not in a financial position to provide the necessary enhancements for training and development for teachers.

A third limitation to my project is the dissemination of a research project that does not yield a reliable or valid tool for the assessment of HOTS and LOTS. While the study I

have reported has collected and analyzed data, the study data did not show that the BRT is not a reliable or valid tool for the assessment of thinking levels as found in student writing. Although this is a limitation of my project, it also brings to light the necessity for the continued exploration of this line of research. One potential reason for this project limitation is that the research study was lacking in comprehensive, job embedded, professional development units specifically intended to familiarize teachers with HOTs and scoring writing for thinking using the BRT. The participants were only briefly exposed to the BRT for scoring HOTs and LOTs in student writing samples. Teachers should be comprehensively trained in designing, implementing, and assessing HOTs (Purnomo, 2017). Further development would be required that was focused solely on using the BRT as a rubric for scoring student writing samples for HOTs.

Recommendations for Alternative Approaches

This project study yielded data that is useful to the field for future studies that will use to make further contributions to the practice of accurately assessing HOTs. The white paper provides current research on the best practices for critical components necessary to fill this gap in practice of assessing students HOTs or LOTs. My overall recommendation is that future research on using the BRT as a rubric would include comprehensive development of teacher's awareness of HOTs and LOTs and the BRT, as well as exhaustive training in using the BRT as a rubric for scoring writing. This study provided only a brief training in the use of the BRT to score writing, while a more in depth training on using of the BRT to score writing could have impacted the statistical significance of the study.

To address the gap in practice of assessing students for the development and demonstration of HOTs using the BRT as a rubric, data were collected to determine if the BRT was a reliable and valid tool. The analysis of the collected data revealed that the BRT was not a reliable or valid scoring tool. There are many factors that may have influenced this data, for example in this study there was only a very brief exposure and training with the BRT as a scoring rubric. An extension to this study that may rectify that deficit could be a more comprehensive training and awareness of HOTs and the BRT as a scoring rubric. This type of enhancement to the existing study could provide the structure needed to reevaluate the BRT to potentially be deemed reliable and valid to fill this gap in practice around assessment of HOTs.

An alternative approach to the project could be the redevelopment of the BRT into a more user friendly rubric for scoring writing. A professional development training center works with a distillation of the BRT which divides it into three levels of cognition by grouping the BRT into three categories; Level 1 or factual information that can be looked up in a book to include the following levels of BRT: remembering, Level 2 or the why or the how which takes the thinking to a procedural level of the BRT: understanding, applying, analyzing, and Level 3 or the universal (human connection) level of conditional knowledge including the *why does this matter* levels of the BRT: evaluating and creating (personal communication, Freeman, 2018). This type of improvement on the structure of the BRT influenced rubric could prove valid and reliable if reevaluated using the same research study methodology.

While additional theoretical frameworks exist to determine which type of thinking is occurring, the specific focus on thinking as visible through writing poses the need for a rubric using language similar to that students would use at each level of thinking. The level of the BRT lends itself to this type of scoring as each level includes action verbs which directly show what type of thinking is being discussed, for example at the highest end of the BRT is creation in which one may identify the verbs: design, compose, hypothesize, collaborate. Perhaps the study could be redefined to assess the influence of certain prompts and the degree to which they elicit language that demonstrates HOTS are occurring. Brookhart (2010) discussed the necessity for educators to design rubrics that represent a balance of content and thinking, which take into account the cognitive intentions for an assignment or prompt. Rather than scoring writing for levels of thinking, the evaluation could fall on the creation of assignments, prompts, or problems that are intended to elicit HOTS and if they are successful or not.

Scholarship, Project Development, and Leadership and Change

Scholarship

Scholarship in a field represents the pursuit of knowledge and academic learning that takes place in the process of deep study. The scholarly pursuit of this project study has prepared me to identify challenges, collect stakeholder input, conduct research, analyze data, and organize it into a scholarly level format to present. This program has also strengthened my skills as an administrator in the areas of engaging in dialogue with my colleagues, and acting in a leadership role to incite positive social change within my

learning community. This program has also helped me to network with other charter school administrators and professionals in the spirit of collaboration within the field.

Throughout my doctoral journey, I have engaged in scholarly research. Writing in a scholarly tone was a challenge at first, but through working with the writing center as well as my committee chair, I enhanced my writing skills. Writing in a scholarly tone requires that one follow the MEAL plan in which you develop a Main idea, provide Evidence, Analysis, and a Lead out. While simple in nature, this format helped me to organize my writing into an acceptable level that was deemed a scholarly tone. The writing center also helped to wean out the passive voice in my writing.

The scholarship required for this doctoral journey also prepared me to think analytically about current research and to synthesis knowledge and apply it to my research. Using the university library was a challenge at first, but through the process of searching for related articles and studies, I honed my skills in locating highly specific information. To complete my literature reviews for my project study, as well as my white paper, I searched and located credible, peer-reviewed articles to support my research topic. Consuming the articles and selecting relevant data to prove I achieved data saturation on my topic required that I read through studies and research to use as evidence of a comprehensive search. This level of exposure to current reliable research provided the next level of awareness of the formatting and scholarly tone used to convey a position within the field of educational research.

An additional challenge I faced throughout this doctoral journey was time management. It was incredibly difficult to work full-time, provide the level of care

necessary for my family, and participate in this doctoral program. Knowing that my health was at risk with atypical migraines induced by stress required that I balance my life to the most minute details. Setting a schedule to work for a few hours every other day after my child's bedtime and my own worked for most of my program. Towards the last year unplanned events in my life threw in new struggles. This program has taught me perseverance, but also the skills and time management necessary to take a project from start to finish no matter what barriers I encounter.

Project Development and Evaluation

This program has helped me develop the skills necessary to develop a quality project to begin to address the gap in practice within my learning community. Based on the data analysis and the necessary next steps for future research to prove the BRT a reliable and valid tool for scoring writing for HOTs, a position paper was the best suited project to pursue. Being that I had limited experience in formatting a white paper, I spent time initially researching white papers related to educational position statements and policy recommendations. Cullen (2018) identified a white paper as an authoritative document used to inform the reader or expert knowledge and research or to argue a specific recommendation or solution within the field.

To begin my white paper, I first identified my audience as stakeholders within my local learning community. As the white paper developed, I decided to include school leaders from two additional charter schools, one of which is within my local school district and the other a part of my charter schools governing body for the state of Colorado. Both charter schools included as an audience for my position paper, focus on

HOTs as an integral component of genuine student learning. At first it was overwhelming to consider consolidating my existing research into easily digestible themes for my white paper. However, once I began this process, new revelations and conjectures began to emerge around themes of best practice in current and necessary next steps for future research in assessment of thinking.

The initial sections of my white paper include background information on the local problem as well as my proposed solution to bridge the gap in practice. I organized the white paper in a fashion that allows the reader to scaffold their understanding of necessary components for the design of learning opportunities during which students can develop HOTs, the implementation of their learning opportunities, as well as the capacities educators must possess to assess thinking using the BRT as a potential rubric. All stakeholders within my learning community, as well as the school leaders from the two charter schools, will receive a link to an evaluation survey via Google Sheets, which they can select to complete as a form of evaluation of my white paper.

Leadership and Change

Throughout my experiences while conducting this study I have developed into a school leader prepared to act as an agent for change within my learning community. It has become habit to uncover the root of problems existing within my learning community and to propose solutions after researching best practice. The information from my research study has been used to take next steps in preparing teachers to understand HOTs, how to design learning opportunities to develop HOTs, how to implement these learning

opportunities and to continue to pursue a reliable and valid rubric to assess student thinking.

In my role as a school leader focused on inciting positive social change within my learning community, I am as a scholar prepared to organize meetings in which the use of current research within the field combined with local data is utilized to address challenges. I have learned to communicate clearly with faculty when sharing a vision or direction, a skill critical to creating the investment of human capital. Additionally, I have developed the habit of creating surveys to gather confidential opinions and ideas from my faculty. For example, surveys similar to the questionnaire accompanying my white paper have been instrumental in getting real time, honest feedback from staff and faculty around certain topics of interest related to improvements in my learning community. This practice has proven incredibly helpful in encouraging the faculty to be heard when they feel passionate about an opportunity or solution.

Reflection on the Importance of the Work

The results of my project study have the potential to incite positive social change which demonstrates the importance of this work. Continued pursuit of the use of the BRT as a rubric for scoring HOTs will contribute to the gap in practice that exists in which HOTs are developed but not assessed. If BRT is used as a rubric, when students are provided with formative feedback demonstrating criteria of HOTs, they will easily be able to adjust their product to add missing criteria based on the rubric. These types of metacognitive, self-reflective habits in a student will provide them the skills to compete in a global job market as graduates. Ganapathy and Wai Kit (2017) asserted that the focus

of traditional school systems is the reproduction of knowledge versus the manipulation and transformation of information that occurs when a student is working in the three upper levels of cognitive skills in the BRT: analysis, synthesis, and evaluation. For there to be necessary change in the national vision regarding these deficits, policymakers must acknowledge the failures of the current system and must make adjustments that match the evolved expectations for students in the 21st century competing for employment.

Current research demonstrates assessment design does not indicate 21st century learning goals including thinking skills (Lamb et al., 2017). Students must demonstrate competencies in critical thinking, problem solving, collaboration, and autonomous independent transfer of knowledge to exercise HOTs (Wagner, 2014). Assessment data focused on thinking and learning must regularly be collected, must inform instruction, and must be pulled from a pool of success criteria universal to the learning community (Brookhart & Chen, 2015; Moss & Brookhart, 2009).

My project will reinforce the necessity for stakeholders and school leaders within my learning community to invest in the pursuit of a reliable and valid rubric using the BRT for scoring student thinking which will ultimately enhance students thinking capacity. I envision the development and assessment of HOTs as a learning benefit that will impact the field of education as it evolves from traditional pursuits of education to new improved learning opportunities in which students thinking is assessed using a reliable and valid assessment tool.

Implications, Applications, and Directions for Future Research

The implications for this project study will affect my local learning community. Research on the challenge of assessing student thinking must be followed up on with further investigation of the BRT as a reliable and valid rubric. To take this research to the next phase, the professional learning community must undergo development in understanding HOTs, designing opportunities for learning in which these HOTs are developed as well as assessed. Within my learning community, we allocate two hours per week for faculty to collaborate and address schoolwide goals. I recommend that the development around HOTs and assessment ensue, which can be followed by a reevaluation of BRT as a reliable and valid tool for assessing thinking.

The involvement of other school leaders and passionate educators will be imperative to the application of this recommendation or the focus on the development of and assessment of HOTs. The digital portfolio software in which the student writing samples are collected and assessed for thinking capacity will provide the ability for tracking student thinking development over time. This type of scaffolded evidence of a student's ability to demonstrate HOTs in writing will provide an alternative form of assessment than previously exists in a currently standardized test heavy field. All stakeholders in the local learning community may benefit if the recommendations of this project are in fact applied.

Conclusion

The project study focused on bridging the gap in practice of assessing student thinking. The participants included eight middle school teachers at the target school.

Participants scored writing samples for thinking using the BRT. Unfortunately, the teachers were only briefly trained to use the BRT as a scoring rubric and the data did not support that it was a reliable or valid scoring tool. However, data variabilities prompted the examination of potential structures and frameworks that may contribute to the redesign and reevaluation of the BRT as a scoring tool. Additionally, variabilities in data from this study may also suggest that a more comprehensive training of the teachers in the understanding and development of HOTs as well as the use of the BRT as a rubric would improve the reliability and validity as the BRT as a scoring tool. Possible recommendations for future research are the redesign and reevaluation of the BRT as a reliable and valid scoring tool following the comprehensive development of educators in teaching HOTs and the concurrent assessment of thinking using the BRT as a scoring rubric. I hope that I will be able to lead my learning community teachers in making these recommendations become a reality.

References

Aagaard, T., & Lund, A. (2013). Mind the gap: Divergent objects of assessment in technology-rich learning environments. *Nordic Journal of Digital Literacy, 8*(4), 225-243. https://doaj.org/article/3478454b64f240d898c243b14851140b

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85(*2), 275-314. doi:10.3102/0034654314551063

Afandi, A., Sajidan, S., Muhammad, A., & Nunuk, S. (2018). Pre-service science teachers' perception about high order thinking skills (HOTS) in the 21st Century. *International Journal of Pedagogy and Teacher Education, 2*(1), 301-308. doi:10.20961/ijpte.v2i1.1825Al-Fayez, M. Q., & Jubran, S. M. (2012). The impact of using the heuristic teaching method on Jordanian mathematics students. *Journal of International Education Research, 8*(4), 453-460. doi:10.19030/jier.v8i4.7293Amer, A. (2006). Reflections on Bloom's revised taxonomy. *Electronic Journal of Research in Educational Psychology*, *4*(1), 213-230. Retrieved from http://www.investigacion-psicopedagogica.com/revista/new/

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.

Angell, K. (2015). The application of reliability and validity measures to assess the effectiveness of an undergraduate citation rubric. *Behavioral and Social Sciences Librarian, 34*(1), 2-15. doi:10.1080/01639269.2015.996481

Argyris, C., & Schon, D. (1974). *Theory in practice*. San Francisco, CA: Jossey-Bass.

Atherton, J. S. (2013). *Learning and teaching; SOLO taxonomy*. Retrieved from http://www.learningandteaching.info/learning/solo.htm

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. *Industrial Psychiatry Journal, 19*(1), 60. https://doi-org.ezp.waldenulibrary.org/10.4103/0972-6748.77642

Baer, J. (2014). *Creativity and divergent thinking: A task-specific approach*. New York, NY: Psychology Press.

Bartell, T. G. (2013). Learning to teach mathematics for social justice: Negotiating social justice and mathematical goals. *Journal for Research in Mathematics Education, 44*(1), 129–163. Retrieved from www.nctm.org

Bauer, J. (2016). *A new approach: Closing the writing gap by using reliable assessment to guide and evaluate cross-curricular argumentative writing* (Thesis). Retrieved from http://dc.uwm.edu/etd/1114/

Beck, S. W. (2006). Subjectivity and intersubjectivity in the teaching and learning of writing. *Research in the Teaching of English*, 413-460. Retrieved from http://www.ncte.org/journals/rte/issues

Benjamin, R., Miller, M. A., Rhodes, T. L., Banta, T. W., Pike, G. R., & Davies, G. (2012). *The seven red herrings about standardized assessments in higher*

education (National Institute for Learning Outcomes Assessment, Occasional Paper No.15). Retrieved from http://www.learningoutcomesassessment.org/occasionalpaperfifteen.htm

Berger, R., Rugen, L., & Woodfin, L. (2014). *Leaders of their own learning: Transforming schools through student-engaged assessment.* San Francisco, CA: Jossey-Bass.

Besser, L. (2011). *Standards and assessment: The core of quality instruction*. L. Besser (Ed.). Englewood, CO: Lead + Learn Press.

Bezuidenhout, M. J., & Alt, H. (2011). Assessment drives learning: Do assessments promote high-level cognitive processing? *South African Journal of Higher Education, 25*(6).

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). Formative and summative assessment: Can they serve learning together. *AERA Chicago, 23*.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, policy & practice*, 7-74. doi: http://dx.doi.org/10.1080/0969595980050102

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364. doi: 10.1007/BF00138871

Biggs, J. B. & Tang, C. (2011). Teaching for Quality Learning at University. (4th Ed.). Maidenhead: McGraw Hill Education & Open University Press

Biggs, J. (2012). Enhancing learning through constructive alignment. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction,*

and learning processes. (pp. 117–136). New York, NY: Cambridge University Press. https://doi-org.ezp.waldenulibrary.org/10.1017/CBO9781139048224.009

Bjornavold, J. (2009). Validation of non-formal and informal learning in Europe: Key developments and challenges. *The Quality of Higher Education*. *6*, 36-71. Retrieved from https://ejournals.vdu.lt/index.php/QHE/issue/archive

Bøhn, H. (2018). Assessing content in a curriculum-based EFL oral exam: The importance of higher-order thinking skills. *Journal Of Language Teaching & Research*, *9*(1), 16-26. doi:10.17507/jltr.0901.03

Brent, D. (2011). Transfer, transformation, and rhetorical knowledge: Insights from transfer theory. *Journal of Business and Technical Communication*, *25*, 396-420. doi: 10.1177/1050651911410951

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. Retrieved from http://www.ascd.org/publications/books/109111.aspx

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343-368.

Brown, C. (2016). *Patterns of Innovation: Showcasing the Nation's Best in 21st Century Learning*. Washington, DC: Pearson Foundation.

Campbell, K. S., & Naidoo, J. S. (2017). Rhetorical Move Structure in High-Tech Marketing White Papers. *Journal Of Business & Technical Communication*, *31*(1), 94-118.

Candal, C. S., & Pioneer Institute for Public Policy Research, C. R. (2015). Massachusetts Charter Public Schools: Best Practices Serving English Language Learners. White Paper No. 140.

Carter, M., Hernandez, A., & Richison, J. (2009). *Interactive notebooks and English language learners: How to scaffold content for academic success*. Portsmouth, NH: Heinemann.

Choudhury, B., Gouldsborough, I., & Shaw, F. L. (2015). The intelligent anatomy spotter: A new approach to incorporate higher levels of Bloom's taxonomy. *Anatomical Sciences Education, 9*, 440-445. doi: 10.1002/ase.1588

Clark, I. (2011). Formative assessment: Policy, perspective, practice. *Florida Journal of Education Administration and Policy, 4*(2), 158-180.

Collet, V. S. (2014). The GIR model: Mentoring for teacher effectiveness. *English Leadership Quarterly, 37*(2), 9-13. Retrieved from http://www.ncte.org/journals/elq/issues/

Colorado Department of Education. (2018). Colorado Measures of Academic Success. Denver: 2018.

Common Core State Standards Initiative. (2013). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Comparative & International Educational Society. (2014). *Renvisioning education for all.* Toronto, Canada.

Confrey, J., & Maloney, A. (2015). A design study of a curriculum and diagnostic assessment system for a learning trajectory on equipartitioning. *ZDM Mathematics Education, 47,* 919-932. doi:10.1007/s11858-015-0699-y

Conley, D. (2015). A new era for educational assessment. *Education Policy Analysis Archives*, *23*(8), 1-36. doi: http://dx.doi.org/10.14507/epaa.v23.1983.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed method approaches*. Los Angeles, CA: SAGE Publications, Inc.

Crockett, R. (2017) Bloom's Digital Taxonomy Verbs. *Global Citizenship.* Retrieved from https://wabisabilearning.com/resources/blooms-digital-taxonomy-verbs/

Crockett, L. W. (2018). LIBRARIANS LEAD THE GROWTH of Information Literacy and Global Digital Citizens. *Knowledge Quest*, 46(4), 28-33.

Cullen, S. (2018). The 12 Best Practices of Contract Management.

Cunningham, C. M., & Lachapelle, C. P. (2014) Designing engineering experiences to engage all students. *Engineering in pre-college settings: Synthesizing research, policy, and practices,* 117-142.

Dagostino, L., Carifio, J., Bauer, J. D., Zhao, Q., & Hashim, N. H. (2015). Using Bloom's Revised Taxonomy to Analyze a Reading Comprehension Instrument. *Current Issues in Education, 18*(2). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1379

Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment and Evaluation in Higher Education,* 42(3).

DeBarger, A., Dornsife, C., Rosier, S., Shechtman, N., & Yarnall, L . (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Retrieved from http://pgbovine.net/OET-Draft-Grit-Report-2-17-13.pdf

Derrington, M. L., & Kirk, J. (2017). Linking job-embedded professional development and mandated teacher evaluation: teacher as learner. *Professional Development in Education*. 43:4, 630-644.

De Swert, K. (2012). Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha. *Center for Politics and Communication*, 1-15.

Dogan, C. D., & Uluman, M. (2017). A Comparison of Rubrics and Graded Category Rating Scales with Various Methods Regarding Raters' Reliability. *Educational Sciences: Theory and Practice*, *17*(2), 631-651.

Dolan, E. L., & Collins, J. P. (2015). We must teach more effectively: Here are four ways to get started. *Molecular Biology of the Cell*, *26*, 2151-2155. doi: 10.1091/mbc.E13-11-0675

Dowd, J. E., Duncan, T., & Reynolds, J. A. (2015). Concept maps for improved science reasoning and writing: Complexity isn't everything. *CBE-Life Sciences Education*, *14*(4), ar39. doi: 10.1187/cbe.15-06-0138

Draper, D. C. (2015). Collaborative instructional strategies to enhance knowledge convergence. *American Journal of Distance Education*, *29*, 109-125. doi: http://dx.doi.org/10.1080/08923647.2015.1023610

Education Excellence Everywhere White Paper. (2016). *Education Journal*, (264), 4-7.

Epstein, S. (1998). *Constructive thinking: The key to emotional intelligence*. Westport, CT: Prager.

Epstein, S., & Meier, P. (1989). Constructive thinking: A broad coping variable with specific components. *Journal of Personality and Social Psychology*, 57, 332-50. 10.1037/0022-3514.57.2.332.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*, 1-4. doi: 10.11648/j.ajtas.20160501.11

Fink, D. L. (2003) *Creating Significant Learning Experiences: An Integrated Approach to Designing*. Hoboken, NJ: Jossey-Bass.

Fisher, D., & Frey, N. (2007). *Checking for understanding: Formative assessment techniques for your classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.

Forehand, M. (2010). Bloom's taxonomy. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology* (pp 41-47). Retrieved from http://www.palieducationsociety.org/images/ebooks%20(13).pdf

Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, *5*(1), 20-33. Retrieved from http://www.ijis.net/

Gallavan, N. P., & Kottler, E. (2012). Advancing social studies learning for the 21st century with divergent thinking. *The Social Studies,* 165-170. doi: http://dx.doi.org/10.1080/00377996.2011.605641

Ganapathy, M., & Wai Kit, L. (2017). Promoting HOTS via ICT in ESL Classrooms. *The Seventh International Language Learning Conference Report.*

Gardner, H. (2010). Five minds for the future. In J. Bellanca and R. Brandt (Eds.), *21st Century Skills: Rethinking How Students Learn* (pp. 9-32). Bloomington, IN: Solution Tree Press.

Gerlach, J. D., & Reinagel, T. P. (2016). Experiential learning in MPA programs: A case for complementarity between internship and service learning requirements. *PS: Political Science & Politics*, 132-138.

Gilboy, M. B., Heinerichs, S., & Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *Journal of Nutrition Education and Behavior*, *47*, 109-114. doi: http://dx.doi.org/10.1016/j.jneb.2014.08.008

Gilmore, J., & Feldon, D. (2010). Measuring graduate students' teaching and research skills through self-report: Descriptive findings and validity evidence. *Online Submission*.

Goldring, E., Rubin, M., Cannata, M., Grissom, J. A., Neumerski, C. M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, *44*(2), 96-104.

Günel, M., Memis, E. K., & Büyükkasap, E. (2010). Effects of the science writing heuristic approach on primary school students' science achievement and attitude toward science course. *Egitim ve Bilim*, 35(155), 49-62. Retrieved from https://www.researchgate.net

Hammill, J., Best, G., & Anderson, J. (2015). Developing student mentor self-regulation skills through formative feedback: Rubric development phase. *Journal of Peer Learning*, 48-58. Retrieved from http://ro.uow.edu.au/ajpl/vol8/iss1/6/

Haolader, F. A., Avi, M. R., & Foysol, K. M. (2015). The taxonomy for learning, teaching and assessing: Current practices at polytechnics in Bangladesh and its effects in developing students' competences. *International Journal for Research in Vocational Education and Training*, 99-118. doi: http://dx.doi.org/10.13152/IJRVET.2.2.9

Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. New York, NY: Teachers College Press.

Harvey, S., & Daniels, H. (2009). *Comprehension & collaboration: Inquiry circles in action*. Portsmouth, NH: Heinemann.

Hassel, H. (2015). Analyzing evidence with rubrics. *Teaching English in the Two Year College, 43*(2), 202-205.

Hassel, H., Klausman, J., Giordano, J. B., O'Rourke, M., Roberts, L., Sullivan, P., & Toth, C. (2015). TYCA white paper on developmental education reforms. Teaching English in the Two-Year College, 42(3), 227-243.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication methods and measures*, *1*(1), 77-89.

Haynes, A., Lisic, E., Goltz, M., Stein, B., & Harris, K. (2016). Moving Beyond Assessment to Improving Students' Critical Thinking Skills: A Model for

Implementing Change. *Journal of The Scholarship Of Teaching & Learning*, *16*(4), 44-61.

Heick, T. (2016). *Teaching thought*. Retrieved October 7, 2016 from http://www.teachthought.com/category/critical-thinking/blooms-taxonomy/.

Hernández, M. L., & Rodríguez, L. F. G. (2016). Encouraging critical thinking development in an EFL classroom through urban legends. *Revista Folios*, *43*, 137-152. Retrieved from http://www.redalyc.org/articulo.oa?id=345943442010

Herman, J., Linn, R., & Moss, F. (2013). *On the road to assessing deeper learning: The status of smarter balanced and PARCC assessment consortia* (CRESST Report 823). Retrieved from http://cresst.org/wp-content/uploads/R823.pdf

Hess, K. K., Jones, B. S., Carlock, D., & Walkup, J. R. (2009). *Cognitive rigor: Blending the strengths of Bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level processes*. Retrieved from ERIC database.

Hohmann, J., & Grillo, M. (2014). Using critical thinking rubrics to increase academic performance, *Journal of College Reading and Learning*, 54:1, 35-51.

Holt, E. A., Young, C., Keetch, J., Larsen, S., & Mollner, B. (2015). The greatest learning return on your pedagogical investment: Alignment, assessment or in-class instruction? *PLOS ONE*, *10*(9), e0137446. doi: http://dx.doi.org/10.1371/journal.pone.0137446

Hong, Y. C., & Choi, I. (2011). Three dimensions of reflective thinking in solving design problems: A conceptual model. *Educational Technology Research and Development*, 687-710. doi: 10.1007/s11423-011-9202-9

Huitt, W. (2011). *A holistic view of education and schooling: Guiding students to develop capacities, acquire virtues, and provide service*. Retrieved from http://www.edpsycinteractive.org/papers/holistic-view-of-schooling-rev.pdf

Hung, W. (2006). The 3C3R model: A conceptual framework for designing problems in PBL. *Interdisciplinary Journal of Problem-based Learning, 1*(1). doi: http://dx.doi.org/10.7771/1541-5015.1006

Hurson, T. (2008). *Think better: An innovator's guide to productive thinking*. New York, New York: McGraw-Hill.

Hyde, C., Stolley, S., & Sakamuno, S. (2015, January 15). *White paper: Purpose and audience*. Retrieved from http://owl.english.purdue.edu/owl/resource

Jailani, & Retnawati, H. (2016). The challenges of junior high school mathematic teachers in implementing the problem-based learning for improving the higher-order thinking skills. *The Online Journal of Counseling and Education, 5*(3), 1–13.

Jaschik, S. (2015). Well-prepared in their own eyes. *Inside Higher Ed, 20*.

Jimerson, J. B., & Childs, J. (2017). Signal and symbol: How state and local policies address data-informed practice. *Educational Policy, 31*(5), 584-614.

Jacobson, L. (2016). GOALS: COHERENCE AND RELEVANCE. *The Learning Professional, (37(6),* 16.

Jones, V. (2012). Essentials for engaged 21st-century learners. *Techniques, 87*(7), 16-19.

Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow* (Vol. 1). New York: Farrar, Straus and Giroux.

Kaufman, J. C., Lee, J., Baer, J., & Lee, S. (2007). Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability. *Thinking Skills And Creativity, 2*(2), 96-106.

Kaiser, L., Kaminski, K., & Foley, J. (2013). *Learning transfer in adult education: New directions for adult and continuing education*. Danvers, MA: Jossey Bass.

Kaldor, E. (2018) *Higher and lower order thinking skills.* Supporting Teaching Excellence at URI.

Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education, 64*, 90-106. doi: 10.1177/0022487112460398

Kleibeuker, S., Koolschijn, P. C., Jolles, D., De Dreu, C., & Crone, E. A. (2013). The neural coding of creative idea generation across adolescence and early adulthood. *Frontiers in Human Neuroscience, 7*, 905. doi: 10.3389/fnhum.2013.00905

Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The technological pedagogical content knowledge framework. In M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 101-111). New York, NY: Springer.

Kolb, D. (1984) *Experiential learning: Experience as the source of learning and development.* Prentice-Hall, Inc., Englewood Cliffs, N.J.

Krippendorff, K. (2011). Computing Krippendorff's alpha-reliability.

Lai, E. R. (2011). *Critical thinking: A literature review*. Retrieved from http://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf

Lam, R. (2017). 'Taking stock of portfolio assessment scholarship: From research to practice'. *Assessing Writing, 31*, 84–97.

Lamb, S., Marie, Q., & Doecke, E. (2017). *Key Skills for the 21st Century: an evidence-based review*. Melbourne, AU: NSW Department of Education.

Lauer, P. A., Christopher, D. E., Firpo-Triplett, R., & Buchting, F. (2014). The impact of short-term professional development on participant outcomes: A review of the literature. *Professional Development In Education, 40*(2), 207-227.

Lo, S. M., Larsen, V. M., & Yee, A. T. (2016). A two-dimensional and non-hierarchical framework of Bloom's taxonomy for biology [Abstract]. *The FASEB Journal, 30*(1 Supplement), 662.14. Retrieved from http://www.fasebj.org/content/30/1_Supplement/662.14

Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice* (2nd ed.). San Francisco, CA: John Wiley & Sons.

Long, C. (2012). Teach your students to fail better with design thinking. *Learning & Leading with Technology, 39*(5), 16-20. Retrieved from http://www.learningandleading-digital.com/learning_leading/201202?pg=1#pg1

Roberts-Mahoney, H., Means, A. J., & Garrison, M. J. (2016). Netflixing human capital development: Personalized learning technology and the corporatization of K-12 education. *Journal Of Education Policy*, *31*(4), 405-420.

McDonald, J. H. (2009) Handbook of biological statistics (Vol.2, pp. 173-181). Baltimore, MD: Sparky House Publishing.

McMillan, J. H. (2013). *Classroom assessment: Pearson new international edition: Principles and practice for effective standards-based instruction* [Kindle version]. Retrieved from Amazon.com

McNamara, C. (2006). *Field guide to consulting and organizational developmental: A collaborative and systems approach to performance, change and learning.* Authenticity Consulting.

McTighe, J., & Wiggins, G. (2013). From common core standards to curriculum: Five big ideas. *The New Hampshire Journal of Education*, 25-31. Retrieved from http://nhascd.net/sites/default/files/pdf/NHJE_pages_2013%20(1).pdf

Meghanathan, N., & He, X. (2016). Correlation and regression analysis for node betweenness centrality. *International Journal of Foundations in Computer Science and Technology*, *6*(6), 1-20.

Meyvis, T., van Osselaer, & Stijn M. J. (2018). Increasing the power of your study by increasing the effect size. *Journal of Consumer Research*, 44, 1157-1173. Available at SSRN: https://ssrn.com/abstract=3065720

Mokharti, K., Yellin, D., Bull, K., & Montgomery, D. (1996). Portfolio assessment in teacher education: Impact on preservice teacher's knowledge and attitudes. *Journal of Teacher Education*, 47 (4), 245-252.

Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Retrieved from https://ebookcentral.proquest.com

Mueller, J. (2016). *Authentic assessment toolbox*. Retrieved January 31, 2017 from http://jfmueller.faculty.noctrl.edu/toolbox/whatisit.htm

Ngang, T. K., Nair, S., & Prachak, B. (2014). Developing instruments to measure thinking skills among Malaysian primary school pupils. *Procedia-Social and Behavorial Sciences*, 116. 3760-3764.

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, 1-36. doi: http://dx.doi.org/10.1080/13546780802490186

Nkhoma, M., Lam, T., Richardson, J., Kam, B., & Lau, K. H. (2016). Developing case-based learning activities based on the revised Bloom's Taxonomy. *Proceedings of Informing Science & IT Education Conference (InSITE) 2016*, 85-93. Retrieved from http://www.informingscience.org/Publications/3496

O'Brien-Moran, M., & Soiferman, L. K. (2010). *How an understanding of cognition and metacognition translates into more effective writing instruction*. Retrieved from https://pdfs.semanticscholar.org/555d/ab62f461b0190c62951abf432c0e3a47b56a.pdf

OCED. (2018). *The future of education and skills education 2030; The future we want*. Organisation for Economic Co-operation and Development, OCED Publishing, Paris.

Panasan, M., & Nuangchalerm, P. (2010). Learning outcomes of project-based and inquiry-based learning activities. *Journal of Social Sciences, 6*(2), 252-255.

Pascal, J., Tíjaro-Rojas, R., Oyander, M. A., & Arce, P. (2017) The acquisition and transfer of knowledge of electrokinetic-hydrodynamics (EKHD) fundamentals: an introductory graduate-level course. *European Journal of Engineering Education, 42*(5), 493-512.

Pearlman, B. (2010). Making 21st century schools: Creating learner-centered school places/workplaces for a new culture of students at work. *Educational Technology*, 14-19.

Pearson. (2010). Partnership for the Assessment of Readiness for Career and College. Riverside, NJ: 2010.

Pecka, S., Schmid, K., & Pozehl, B. (2014). Psychometric testing of the Pecka grading rubric for evaluating higher-order thinking in distance learning. *AANA Journal, 82*(6), 449–456. Retrieved from https://ezp.waldenulibrary.org/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=103918145&site=eds-live&scope=site

Peen, T. Y., & Arshad, M. Y. (2014). Teacher and student questions: A case study in Malaysian secondary school problem-based learning. *Asian Social Science, 10*(4), 174. doi: http://dx.doi.org/10.5539/ass.v10n4p174

Phelan, C., & Wren, J. (2006). *Exploring reliability in academic assessment*. Retrieved from https://www.uni.edu/chfasoa/reliabilityandvalidity.htm.

Pretorius, L., van Mourik, G. P., & Barratt, C. (2017). Student choice and higher-order thinking: Using a novel flexible assessment regime combined with critical thinking activities to encourage the development of higher order thinking. *International Journal Of Teaching & Learning In Higher Education*, *29*(2), 389-401.

Purnomo, Y. W. (2017). The complex relationship between teachers' mathematics-related beliefs and their practices in mathematics class. *New Educational Review, 47*(1), 200–210. doi:10.15804/ tner.2017.47.1.16.

Radmehr, F., & Drake, M. (2017). Revised Bloom's taxonomy and integral calculus: unpacking the knowledge dimension. *International Journal Of Mathematical Education In Science & Technology*, *48*(8), 1206-1224.

Raiyn, J., & Tilchin, O. (2016). The impact of adaptive complex assessment on the HOT skill development of students. *World Journal Of Education*, *6*(2), 12-19.

Rashid, G. J., & Duys, D. K. (2015). Counselor cognitive complexity: Correlating and comparing the Myers–Briggs Type Indicator with the Role Category Questionnaire. *Journal of Employment Counseling*, *52*(2), 77-86.

Rembach, L., & Dison, L. (2016). Transforming taxonomies into rubrics: Using SOLO in social science and inclusive education. *Perspectives in Education*, *34*(1), 68-83.

Renaissance. (2019). Standardized Test for the Assessment of Reading. Wisconsin: 2019.

Retna, K. S., & Ng, P. T. (2016). The application of learning organization to enhance learning in Singapore schools. *Management in Education*, *30*(1), 10-18. doi: 10.1177/0892020615619665.

Retnawati, H. H., Djidu, H. H., Kartianoml, K., Apino, E. A., & Anazifa, R. R. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems Of Education In The 21St Century*, *76*(2), 215-230.

Rojas, S. L., & Widiger, T. A. (2014). Convergent and discriminant validity of the Five Factor Form. *Assessment, 21*(2), 143-157.

Rothstein, D., & Santana, L. (2011). *Make just one change: Teach students to ask their own questions*. Cambridge, MA: Harvard Education Press.

Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice* (2nd ed.). Amsterdam: Academic Press, an imprint of Elsevier.

Sabourin, K., Lowe, B., & Bowman, J. (2015). The use of ePortfolios to support first year student's metacognitive thinking. *Educational Technology Publications*, Paper 2. Retrieved from http://fisherpub.sjfc.edu/edtech_pub/2/

Scott, L. A. (2017). 21st century skills early learning framework. Partnership for 21st Century Skill (P21). Retrieved from http://www.p21.org/storage/documents/EarlyLearning_Framework/ P21_ELF_ Framework_Final.pdf.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researche*r, 29, 4-14.

Siegel, D. J. (2010). *Mindsight: The new science of personal transformation.* New York, NY: Bantam Books.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 68-85. doi: http://dx.doi.org/10.1037/1931-3896.2.2.68

Soland, J., Hamilton, L., & Stecher, B. (2013). *Measuring 21st Century Competencies: Guidance for Educators*. Global Cities Education Networ: Rand Corporation.

Sotiriou, S., Riviou, K., Cherouvis, S., Chelioti, E., & Bogner, F. f. (2016). Introducing Large-Scale Innovation in Schools. *Journal Of Science Education & Technology*, *25*(4), 541-549.

Sotiriadou, P., & Hill, B. (2015). Using scaffolding to promote sport management graduates' critical thinking. *Annals of Leisure Research*, 105-122. doi: http://dx.doi.org/10.1080/11745398.2014.925406

Spruce, R., & Bol, L. (2015). Teacher beliefs, knowledge, and practice of self-regulated learning. *Metacognition and Learning, 10*(2), 245-277.

Steedle, J., & Ferrara, S. (2016). Evaluating comparative judgement as an approach to essay scoring. *Applied Measurement in Education*, *29*(3), 211-223.

Sultana, F. (2010). *An initial study of a method for instructing educators about the revised taxonomy*. Retrieved from ERIC database. (ED523761)

Taylor, L., & Galaczi, E. (2011). *Scoring validity. Studies in language testing. Examining Speaking. Research and Practice in Assessing Second Language Speaking*. Cambridge: Cambridge University Press.

Thompson, R., & O' Loughlin, D. (2015). The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anatomical Sciences Education, 8*, 493-501. doi: 10.1002/ase.1507

Thompson, E., Luxton-Reilly, A., Whalley, J. L., Hu, M., & Robbins, P. (2008). *Bloom's taxonomy for CS assessment*. Retrieved from http://crpit.com/confpapers/CRPITV78Thompson.pdf

Tidwell, A. L. (2015). Assessing student problem solving using structured versus unstructured case analysis. *Academy of Educational Leadership Journal, 19*(3), 309-324.

Tíjaro-Rojas, R., Arce-Trigatti, A., Cupp, J., Pascal, J., & Arce, P. E. (2016). A systematic and integrative sequence approach (SISA) for mastery learning: Anchoring Bloom's Revised Taxonomy to student learning. *Education for Chemical Engineers, 17*, 31-43. doi: http://dx.doi.org/10.1016/j.ece.2016.06.001

Trochinm, W. M. K. (2006). Convergent and discriminant validity. *Research Methods Knowledge Base*, 87(5), 875.

Vanlommel, K., & Schildkamp, K. (2018). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal*. https://doi.org/10.3102/0002831218803891

Vandal, B. (2012). *Riding the storm out: Developmental education reform as a key component of Common Core implementation*. Retrieved from ERIC database. (ED541979)

Vasan, M., Venkatachary, R., & Freebody, P. (2006). Can collaboration and self-direction be learned? A procedural framework for problem-based learning. *Planning and Changing*, *37*(1/2), 24-37.

Vista, A., Care, E., & Griffin, P. (2015). A new approach towards marking large-scale complex assessments: Developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. *Assessing Writing*, *24*, 1-15. doi: http://dx.doi.org/10.1016/j.asw.2014.11.001

Voogt, J., & Roblin, N. (2012) A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies, *Journal of Curriculum Studies*, *44*(3), 299-321, DOI: 10.1080/00220272.2012.668938

Wagner, T. (2014). *The global achievement gap* (Upd. Rev. Ed.). New York, NY: Basic Books.

Wason, H. R., Sinvhal, A., & Bhattacharya, B. (2016). Developing suitable pedagogical methods for outcome-based e-learning. In *Teaching and Learning with Technology: Proceedings of the 2015 Global Conference on Teaching and Learning with Technology (CTLT)*, (59-70). Retrieved from http://www.worldscientific.com/worldscibooks/10.1142/9880#t=toc

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Wilson, L. O. (2013). Anderson and Krathwohl-Understanding the New Version of Bloom's Taxonomy. *The Second Principle. The work of Leslie Owen Wilson. Ed. D.*

Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, *45*, 128–136. https://doi-org.ezp.waldenulibrary.org/10.1016/j.tate.2014.10.002

Yassin, S. M., Tek, O.E., Alimon, H., Baharom, S., & Ying, L. Y. (2010). Teaching science through English: Engaging pupils cognitively. *International CLIL research journal*, 1(3), 45-59.

Yildiz, L. M. (2011). English VG1 level oral examinations: How are they designed, conducted and assessed? (Unpublished). University of Oslo, Oslo, Norway.

Yoruk, S., & Runco, M. A. (2014). The neuroscience of divergent thinking. *Activitas Nervosa Superior*, 1-16.

Young, K., James, K., & Noy, S. (2016). Exploration of a Reflective Practice Rubric. *Asia-Pacific Journal Of Cooperative Education, 17*(2), 135-147.

Zull, J. E. (2006). Key aspects of how the brain learns. *New directions for Adult and Continuing Education, 2006*(110), 3-9. doi: 10.1002/ace.213

Zull, J. E. (2011). *From brain to mind: Using neuroscience to guide change in education*. Sterling, VA: Stylus Publishing, LLC

Appendix A: White Paper

Developing and Evaluating High Order Thinking Skills

Abstract

The purpose of this study was to determine if Bloom's Revised Taxonomy (BRT) could be used as a rubric to validate and reliably assess student thinking as demonstrated in written work. This quantitative, non-experimental project study is rooted in Bloom's developmental theory of knowledge construction through varying levels of thinking skills. This study explored the inter-rater reliability of a scoring BRT rubric for assessing students' levels of lower to higher order thinking. This study promotes positive social change validating a rubric to quantify and assess student thinking. This type of structured scoring process could lead to more widespread teaching of and assessing higher order thinking skills (HOTs) that promote quality of life in the 21st century.

Introduction

Students are ill-prepared to face uncertainty and develop solutions for 21st century challenges; many of which are still unidentified. In order to be prepared, students must learn higher-order thinking skills (HOTS) (Heick, 2016). There are many initiatives to teach HOTS. There are some assessments for HOTS, but they are unwieldy and the scoring is subjective (Silvia et al., 2008). In order to remedy this deficit in our society, educators must begin to assess student's ability to think.

Currently, the United States of America, a leader in developed nations, lags in comparison to other nations in achievement as represented on international assessment charts (Comparative & International Education Society, 2014). A contributing factor to

this deficit in achievement is that teacher do not have a reliable tool for assessing student writing. Bloom's Revised Taxonomy is a quantitative categorical scoring taxonomy that could be used to measure HOTS. This could have a great positive social impact as some have said that our students' HOTS may build a bridge to close the achievement gap between the US Education System and our global competitors (Wiliam, 2011).

The Study

A study was conducted to examine the reliability and validity of using the BRT to score student writing for varied levels of thinking. In the design of this study the author intended to have teachers score writing samples using the BRT as a taxonomy for scoring. The study included a brief training session preceding the scoring of the student writing samples in which the author provided an overview of the BRT and some sample exercises in scoring writing. The results of the study could not prove that the BRT was a reliable and valid scoring tool for student thinking. The mistake made was in that the design only included a brief participant training. If intensive training with using the BRT to score student work was implemented, the BRT could very well be the solution needed to the problem of scoring writing for varied levels of thinking, including HOTs. After the research study was complete, the findings were used to design a project that would prompt positive social change within my learning community. For this project, I selected the dissemination of a position paper.

The goals of the position paper are to discuss the lack of assessment tools for scoring students levels of thinking. An explanation of the suggested adjustments will be addressed to prompt further research on this deficiency within the field of education.

In my research, I used a quantitative research design. The research questions asked in this study are:

RQ1: What is the relationship between teacher grades for each writing sample and the BRT-based LOTS-HOTS ratings for each writing sample?

RQ2. Will there be moderate (>.7) or better inter-rater reliability demonstrated by middle school teachers' ratings using BRT for scoring multiple writing samples of student demonstrations of thinking?

RQ3. Will there be moderate (>.7) or better inter-rater reliability demonstrated by the *true population* Krippendorf alpha estimates between middle school teachers' ratings using the BRT scoring rubric for multiple samples of student demonstrations of thinking?

I collected data using Google Sheets to confidentially collect my data from the teacher participants. I organized the data collected using the student writing assignments by copying and pasting the data from Google Sheets into predetermined variables defined in the IBM Statistical Package for the Social Sciences (SPSS) to determine relationships between original teacher scores and the BRT rubric scores (validity) and inter-rater reliability. My data analysis did not find evidence of validity or reliability that the BRT rubric in its current form was a valid and reliable rubric for assessing levels of thinking through the analysis of student writing samples.

The following graphic in Table 1: Levels of Bloom's Revised Taxonomy was used as the rubric structure for scoring writing in this study in combination with the graphic in Table 2: Bloom's Revised Taxonomy 249 Verbs.

Table 1

Levels of Bloom's Revised Taxonomy

6	Creating
5	Evaluating
4	Analyzing
3	Applying
2	Understanding
1	Remembering

Note. From "Bloom's Revised Taxonomy (BRT) Verbs and Similar Verbs from Higher Order Thinking Skills (HOTS) down to Lower Order Thinking Skills (LOTS)," by *Crockett, R.* Staff Global Citizenship Copyright 2017. Adapted with permission.

Table 2

Bloom's Revised Taxonomy 249 Action Verbs

Knowledge	Understand	Apply	Analyze	Evaluate	Create
define	explain	solve	compare	reframe	design
identify	describe	apply	classify	criticize	compose
describe	interpret	illustrate	contrast	evaluate	create
label	paraphrase	modify	distinguish	order	plan
list	summarize	use	infer	appraise	combine
name	classify	calculate	separate	judge	formulate
state	compare	change	explain	support	invent
match	differentiate	choose	select	compare	hypothesize
recognize	discuss	demonstrate	categorize	decide	sustitute
select	distinguish	discover	connect	discriminate	write
examine	extend	experiment	differentiate	recommend	compile
locate	predict	relate	discriminate	summarize	construct
memorize	associate	show	divide	assess	develop
quote	contrast	sketch	order	choose	generalize
recall	convert	complete	point out	convince	integrate
reproduce	demonstrate	construct	prioritize	defend	modify
tabulate	estimate	dramatize	subdivide	estimate	organize
tell	express	interpret	survey	find errors	prepare
copy	identify	manipulate	advertise	grade	produce
discover	indicate	paint	appraise	measure	rearrange
duplicate	infer	prepare	break down	predict	rewrite
enumerate	relate	produce	calculate	rank	role-play
listen	restate	report	conclude	score	adapt
observe	select	teach	correlate	select	anticipate
omit	translate	act	criticize	test	arrange
read	ask	administer	deduce	argue	assemble
recite	cite	articulate	devise	conclude	choose
record	discover	chart	diagram	consider	collaborate
repeat	generalize	collect	dissect	critique	collect
retell	give examples	compute	estimate	debate	devise
visualize	group	determine	evaluate	distinguish	express
	illustrate	develop	experiment	editorialize	facilitate
	judge	employ	focus	justify	imagine
	observe	establish	illustrate	persuade	infer
	order	examine	organize	rate	intervene
	report	explain	outline	weigh	justify
	represent	interview	plan		make
	research	judge	question		manage
	review	list	test		negotiate
	rewrite	operate			originate
	show	practice			propose
	trace	predict			reorganize
	transform	record			report
		schedule			revise
		simulate			schematize
		transfer			simulate
		write			solve
					speculate
					structure
					support
					test
					validate

Note. From "249 Bloom's Taxonomy Verbs for Critical Thinking," by Heick, T. Teach Thought Staff. Copyright 2017 by TeachThought. Adapted with permission.

Bloom's Revised Taxonomy

Among the constructivist learning theories, there are social constructivist theories and cognitive constructivist theories (Biggs, 1996). Constructivist learning is an active learning process through which learners scaffold and adapt what they know according to new information (Shepard, 2000). Within constructivist learning theory there are two main assessment frameworks; a) authentic assessments which focus on higher order thinking and knowledge integration, and b) developmental assessments which focus on diagnosing a student's readiness in order to adjust instruction (Mokharti, Yellin, Bull, & Montgomery, 1996).

BRT is based upon Benjamin Bloom's taxonomy, originally designed by Bloom in 1956 along with a group of educational psychologists, classified educational objectives into six categories (Sultana, 2010). After more than forty years of instructional design based on Bloom's original taxonomy, Anderson, Krathwohl, and Bloom (2001) revised the taxonomy to include the previously classified thinking skills as cognitive strategies in verb form with create replacing evaluate at the top of the hierarchy. BRT is a widely used guide for the design of curriculum and evaluation of instructional opportunities within the field (Forehand, 2010; Thompson & O' Loughlin, 2015).

Table 1

Bloom's Revised Taxonomy (BRT) Verbs and Similar Verbs from Higher Order Thinking Skills (HOTS) down to Lower Order Thinking Skills (LOTS)

	HOTS: Higher Order Thinking Skills
BRT Verbs	**Similar Verbs**
Creating	Designing, constructing, planning, producing, inventing, inventing, devising, making
Evaluating	Checking, hypothesizing, critiquing, experimenting, judging, testing, detecting
Analyzing	Comparing, organizing, deconstructing, attributing, outlining, finding, structuring, integrating
Applying	Implementing, carrying out, using, executing
Understanding	Interpreting, summarizing, inferring, paraphrasing, classifying, comparing, explaining, exemplifying
Remembering	Recognizing, listing, describing, identifying, retrieving, naming, locating, finding
	LOTS: Lower Order Thinking Skills

Note. From "Bloom's Revised Taxonomy (BRT) Verbs and Similar Verbs from Higher Order Thinking Skills (HOTS) down to Lower Order Thinking Skills (LOTS)," by Crockett, R. Global Citizen Copyright 2017. Adapted with permission.

While changes have occurred in the approach to teaching, there is still a gap in the practice of developing and implementing assessments which require students to demonstrate higher order cognitive progressions including the BRT categories of evaluation and creation of new knowledge, as well as metacognitive awareness of these thinking skills (Draper, 2015). Bezuidenhout and Alt (2011) noted that Bloom's Revised Taxonomy was developed to foster the development of assessments focused on varied cognitive demonstrations (Bezuidenhout & Alt, 2011; Adams, 2015). Haolader, Avi and Foysol (2015) identify that this type of structured construction of knowledge occurs in the design phase of education. For example, BRT is used to design questions to ask students during small group discussions at a particular level such as the understanding level. Haolader et al. (2015) point out that BRT is rarely, if ever part of the design of assessment tools. This study seeks to use BRT explicitly for assessment as a rubric.

Indeed, most educators currently practicing in the field do not commonly assess BRT levels at any point. Instead, teachers' assessments largely focus on summative assessment of content recall and organization (Huitt, 2011). Educators could emphasize that instead of task completion, that the ultimate goal is profound and genuine learning. Instead of just a grade, we could have an evaluation of whether or not thinking and learning have taken place (Brookhart, 2013). Teachers could have a consistent focus on student thinking assessment with BRT rubrics; teachers could use BRT rubric data as a central tool for driving the next instructional steps for all students (Wiliam, 2011).

Educators strive to stimulate higher levels of thinking through learning opportunities, therefore the assessment of student progress is required for continued

growth (Cunningham & De Aquino, 2015). To teach for advancements in student thinking, we must be able to assess student thinking as our students develop their ability to apply content effectively (Nkhoma, Lam, Sriratanaviriyakul, Richardson, Kam, & Lau, 2017). The field requires the development of a quantitative assessment of thinking to track this growth and evaluate student preparedness to tackle tasks that require higher-order thinking (Rembach & Dison, 2016). The BRT rubric, with additional research and development, could one day serve that purpose.

Types of Thinking

Thinking is constructed in a context. Much like instructional strategies vary based on the students in a given classroom, the type of thinking one employs depends on the application of thought required (Hung, 2006). Different types of thinking are good for different types of tasks. In the event that a task requires divergent thinking, the thinker would generate as many possible solutions or theories as one can regard a concept or topic (Gallavan & Kottler, 2012; Kaufman, Lee, Baer, & Lee, 2007). Hurson (2008) described productive thinking as a process through which one combines knowledge with critical or creative thinking. My analysis of literature reveals a gap in practice on assessing students for the critical capacity of either divergent or productive thinking (Lam, 2017).

This gap is highlighted by the historical emphasis in curriculum and standardized assessments on convergent thinking tasks, or tasks in which thinkers are expected to apply content or knowledge to complete a finite or defined task. There is a lack of instruction in using converged ideas or content associations to create diverging solutions

to proposed challenges (Kaufman et al., 2007). Recent attempts have been made to infiltrate standardized assessments with performance tasks which require varied levels of divergent thinking (CCSSI, 2010). There appears to be a disconnect between the convergent thinking required on tests and the divergent thinking required for solving real-world problems.

Tests and real-world problems both have objectives. Governments and schools list objectives in standards, and then assess based upon those standards. While students are completing tasks in school, the idea is that they learn and develop an understanding of a concept or skill often tied to a standard or benchmark. However, a student may arrive at an answer being unsure of how they got to the answer, because subconscious connections were being made by their mind all the while they were working on a task (Runco, 2014). That is, we rarely can see a person's thinking but rather simply a finished product that reflects the scaffolded thinking used to create the product (Sotiriadou & Hill, 2015). It may be useful to have assessments of varied levels and applications of thinking. They would provide a gauge for educators, and be useful in the practice of prompting learning (Harvey & Daniels, 2009; Tíjaro-Rojas et al., 2016).

Runco (2014) demonstrated that it is through subconscious associations that learners shift their level of understanding, while thinking, from superficial representations (content knowledge) to complex representations and transfer. Once one has reached the more complex levels of thinking, genuine and lasting learning has occurred and independent transfer is possible in new and unknown situations ((Dagostino, Carifio, Bauer, Zhao, & Hashim, 2015). In this same vein of learning Argyris and Schon (1974)

identified single and double loop learning as components of their theory of action in which human beings are agents of change. Single loop learning identifies one's decision to follow existing rules, while double loop learning (representative of middle levels of thinking in BRT) occurs when one adapts their thinking and generates ideas about the existing rules (Argyris & Schon, 1974). Triple loop learning takes thinking to the highest level of BRT (create) and occurs when on creates new rules based on what they have learned about a certain topic or situation (McNamara, 2006). This notion of transfer, of taking knowledge and applying it, is important because it is the ultimate assessment; do students use what they learn in situations outside of classrooms?

Transfer occurs when prior learning influences future performance (Clark, 2011). Varied levels of transfer have been noted: near, far, and further transfer. The degree is based on the connection and similarities between the knowledge and the situation in which one is trying to perform a task that requires that knowledge (Kaiser, Kaminski, & Foley, 2013). Brent (2011) asserted that after the transfer of knowledge has occurred, the new resulting knowledge has been transformed – the knowledge is now associated with the situation in which is successfully helped solve the problem. A classroom focused on problem-based learning offers potential to observe stages of knowledge incorporation, transfer, and transformation through various instructional strategies (Panasan & Nuangchalerm, 2010; Tidwell, 2015). Having a learning environment and educator designing opportunities for transfer is helpful.

The Experiential Learning Theory popularized by Kolb (1984) identified a four-cycle learning process in which once associates concrete-abstract and reflective-active

dimensions of learning. This cycle of learning begins with an experience, followed by an assimilation of the new knowledge with old values to be reflected on and transferred from abstract thoughts to concrete associations (Kolb, 1984). To further understand the critical nature of transfer as an ultimate test for learning, we can examine the biological aspects of the physical learning process. Zull (2011), drawing from the prior works of Kolb's Experiential Learning Theory, proposes that the brain physically changes as one learns though the process of what he call the Four Pillars: Gathering Information, Reflection, Creating, and Testing. Throughout this process Zull (2006) noted that in the early phases of learning one gathers data through sensory inputs and assigns a value to each gathered data point. The process through which the data moves from the sensory neocortex to the association regions Zull (2006) labels as the reflection phase. This is followed by the creation phase in which these new associations engage working memory to create new ideas or theories. The final pillar of testing engages the motor brain to transfer of the created theory from abstract to concrete through application to a new situation or challenge (Zull, 2011).

Indeed, according to Nokes (2009), there is a need for educators to provide learning experiences that include the meta-cognitive practice of transfer. Students must be explicitly taught how to become cognizant of opportunities in which they may transfer knowledge and have the skills to proceed. Ultimately, steps in the instructional process are needed during which transfer skills are explicitly taught and transfer itself is measured (Nokes, 2009). Gardner (2010) noted the importance of educational opportunities focused on the transformation of knowledge which foster growth in

citizenship and the development of social involvement. For students to reach their full potential in terms of transfer, research shows that they should be intrinsically motivated and acting on their volition. It is not enough to have instruction; students' emotional state must be figured into the learning equation (Zull, 2006).

In the design of learning opportunities teachers could benefit from the implementation of a reliable, valid rubric for the assessment of student development of thinking capacity from the lowest observable levels to the highest observable or documentable levels. The BRT could be organized in a way that would potentially contribute to this type of scoring rubric as varied types of thinking can be categorized using verbs in each level of the BRT. Continuing to investigate the reliability and validity of current frameworks and processes for learning as we all as ways that students thinking capacity can be documented would be a viable next step in the identification of such a generalizable tool.

Learning Environment and Student Engagement

Thinking is an internal process that we cannot see, so we must depend on models and research documenting best practices to encourage students to develop and practice higher quality thinking. In the process of learning, students filter through their personal knowledge base, experiences, and internal reactions. Through sound instructional practice students process new associations and genuine, transferable learning occurs (Spruce & Bol, 2015). Ultimately, based on the learned ability to think well, good thinkers develop original ideas and thoughts to help them solve future challenges (Halpern, 2013). Students must demonstrate thinking mastery on assessments and other learning tasks

regularly as designed, implemented, and monitored by a skilled educator (Tíjaro-Rojas, Arce-Trigatti, Pascal, & Arce, 2016).

Documented evidence of students thinking as they progress towards learning should occur in a formative, ongoing thread focused on providing a gauge for educators and students through a multifaceted reflective cycle (Brookhart, 2013). The path of learning winds in many directions, and assessment is a necessary feedback loop to stay on the course toward the instructional objective (Haynes, Lisic, Goltz, Stein, & Harris, 2016). Collecting, organizing, and maintaining the scoring of writing assessments has been a long-standing challenge, but with new technology there are new opportunities as discussed below (Conley, 2015).

A project-based or problem based learning portfolio approach provides students with a software platform to complete activities and associated writing samples at various phases throughout a project timeline, thus providing necessary evidence of student thinking and learning. A digital portfolio system to track the problem based learning process, organizationally supports tracking and assessment of students development of thinking ability thus fostering the creation of new knowledge out of existing information (Fink, 2003). Educators can then score students' work using a cognitive progression such as BRT to provide objective, structured feedback to track thinking through the varied demonstrations collected in a digital portfolio system.

Further examination of the development of an assessment tool to evaluate students HOTS an LOTS combined with a focus on learning environment and student engagement could guide educators towards the creation of enhanced, authentic learning

opportunities relevant to the skills necessary to complete in a global economy. Soft skills such as listening, collaboration, problem solving and reflection are highly relevant 21st Century Skills.

21st Century Learning

For lasting changes to occur in education, it is imperative that policymakers, administrators, and most importantly practitioners recognize necessary changes in learner expectations as well as the purpose of teaching; teaching students to think (Retna & Ng, 2016). The initial shift requires the transition from teacher as keeper of knowledge to the teacher in the role of facilitator and guide (Dolan & Collins, 2015; Shepard, 2000). Collet (2014) emphasized a balance between self-direction and expert mentoring as the key to successful learning. A key component to fostering genuine learning is learner participation with a classroom dynamic rooted in the value of developing thinking skills. Students in a reflexive and thought-based classroom are likely to own their learning processes, and emphasize questioning as a method of learning (Peen & Arshad, 2014). Student development of questioning techniques provides a method for motivating and engaging students in authentic concerns that they may have or passions they chose to pursue while promoting collaborative dialogue and other necessary 21st century skills (Rothstein & Santana, 2011). Beyond focusing on content as the only objective, the development of quality thinking is a higher educational objective and goal (Choudhury, Gouldsborough, & Shaw, 2015).

To engage students in 21st-century habits of learning content and thinking, the design and implementation of a problem-based learning program offers a combination of

the elements more supportive than traditional spoon-feeding of information. Within science classrooms, the heuristic inquiry approach is used to learn concepts and skills within the domain (Günel, Memis, & Büyükkasap, 2010; Lo, Larsen, & Yee, 2016). A heuristic learned through discovery or inquiry provides improved understanding, increased connections, and an increase in cognitive activity (Al-Fayez & Jubran, 2012). In short, heuristics are common ways of thinking that can be applied, or transferred to new situations.

Therefore, we need learning environments in which an educator designs opportunities for students to engage in the active discovery of methods and heuristics of thinking (Bezuidenhout & Alt, 2011). This type of learning design requires the use of processes and instruments (such as the BRT rubric) for gauging thinking and student growth towards independent near transfer (Anderson et al., 2001). Hong and Choi (2011) examined the relationships and patterns occurring during reflective thinking of novice to expertise in a field. Hong and Choi (2011) was working on developing a research-based learning progression that students travel from novice to expert. The BRT rubric is a more general progression from the novice level of remembering to the most expert level of creating.

HOTs and LOTs

The development of HOTs is essential for students to reach their potential to become effective, contributing members of society as adults. Developing the capacity to solve everyday problems and establish solutions when faced with a challenge is not something that is currently taught in traditional school systems. Traditional schooling

models primarily utilize the bottom levels of the BRT and fail to bridge the gap between the concepts and content learned and the HOTs necessary to use them (Kaldor, 2018). Scott (2017) delineates three main frameworks of 21st-century skills: 1) learning and innovation skills, 2) life and career skills, and 3) information, media, and technology skills of which HOTs are grouped under the learning and innovation skills. Additionally, Ganapathy & Wai Kit (2017) supports that the focus of traditional school systems is the reproduction of knowledge versus the manipulation transformation of information that occurs when a student is working in the three upper levels of cognitive skills in the BRT: analysis, synthesis, and evaluation. In order for there to be necessary change in the national vision regarding these deficits, policymakers must acknowledge the failures of the current system and must make adjustments that match the evolved expectations for students in the 21st century competing for employment.

Recent research on the development and assessment of HOTs proposes the engagement of students in their learning in active learning and student-centered ways. Retnawati, Djidu, Kartianoml, Apino, and Anazifa (2018) discuss the element of synergy between stakeholders in the pursuit of training teachers to train students in HOTs. To effectively implement the development of HOTs, teachers, curricular updates, and the continued development of teaching professionals must be at the forefront of this critical implementation (Purnomo, 2017). Educators must consider current research and best practice to guide the development of authentic learning opportunities.

Problem based learning, discovery learning, inquiry based learning, and any model using contextualized problems will provide the necessary training experiences for

students through which they can develop HOTs (Gerard, Kihyun, McElhaney, Liu, Rafferty & Linn, 2016; Retnawati et al., 2018). Bartell (2013) proposes that teachers can achieve these types of experiences within their practice by playing an active role in planning, implementing, and evaluating HOTs oriented learning. A challenge in the implementation of HOTs based learning experiences is the misunderstandings that teacher generally have around the types of learning opportunities that could be used to train students for HOTs.

While teachers generally value HOTs as the skills students need to solve everyday problems, they are unable to articulate the steps of operational implementation of the necessary learning experiences (Jailani & Retnawati, 2016). According to Jailani and Retnawati (2016) teachers have identified methods for the assessment of HOTs such as contextual based essay prompts, but have not found the link between the measurement of HOTs using the BRT. They note HOTs as the top three categories: analysis, synthesis, and evaluation (Jailani & Retnawati, 2016). To move forward, educators must develop a clear understanding of HOTs and how to develop, implement, and assess HOTs in order to train students adequately.

Assessment

The absence of generalizable assessments that measure student thinking capacity is the problem within the field of education. On a large scale, assessment design does not indicate 21st century learning goals including thinking skills (Brown, 2016). Students must demonstrate competencies in critical thinking, problem solving, collaboration, and autonomous independent transfer of knowledge to exercise HOTs (Wagner, 2014).

Assessment data focused on thinking and learning must regularly be collected, must inform instruction, and must be pulled from a pool of success criteria universal to the learning community (Brookhart & Chen, 2015; Moss & Brookhart, 2009).

Fortunately, based on the deficits businesses and colleges have identified, the assessment of thinking has begun to surface in state standardized assessments. Such standardized assessment tools include PARCC and Smarter Balanced Assessments in the form of Performance-Based Assessment components (Benjamin et al., 2012; Herman, Linn, & Moss, 2013). Considering that large-scale testing corporations' attention has begun to focus on this area of need, policy makers and stakeholders informing practice in the field have started to take notice (Cunningham & De Aquino, 2015).

This subjectivity inherent in assessment of written work manifests through practice in many ways. Hess, Jones, Carlock & Walkup (2009) noted the discrepancies in teacher scoring as they fall into old habits of scoring on academic enablers such as student past behaviors or achievements. For example, if a teacher scores an essay and the rubric is vague they are likely to factor in historical subjective observations and associations from interactions with the student in the past. Additionally, a student's actual academic competence and habits may factor into what a teacher identifies or focuses on if, for example, the student is frequently late in turning in work. The attitude of the teacher may be less open to possibilities of the range of LOTS to HOTS in each student's writing.

While a good deal of research is available related to classroom assessment, there is a gap in research around the documentation of student LOTS and HOTS (Wiggins &

McTighe, 1998). McMillan (2013) identified the need to develop principles of assessment that document student learning, addressing specifically the necessity for a supporting body of research on classroom assessment. Furthermore, Moss & Brookhart (2009) discuss the need for developing evidence of in-depth descriptions of how teachers summarize and document learning and how that learning progresses. In the 21st century, with the transformation in the contexts for assessment, Aagaard and Lund (2013) identified the lack of educator's experience in how to assess collaborative and interactively constructed learning (p. 223).

There is a difference between design of learning questions and the assessment of the thinking generated from those questions. While there are structures in place, such as BRT, to guide the design of learning opportunities and questions that address higher order thinking, there is a breakdown in the assessment of the responses to the questions focused on higher order thinking (Vista, Care & Griffin, 2015). Bøhn (2018) discussed his research in which teachers are familiar with the assessment of the what (knowledge) but are unfamiliar with the how (cognition) which calls for the further development teachers to understand this difference and begin to develop assessment tools that evaluate student's abilities to present their discoveries.

In the design of evaluation tools to use in a formative or summative manner, indicators must be identified for ideal student outcomes for specific tasks (ideally HOTs based opportunities) within a given discipline. For example, Atherton (2013) discuss the phases of learning using a Structure of Observed Learning Outcomes (SOLO) taxonomy, in which indicators are checked off as the students' learning progresses through Piagetian

developmental phases beginning with the pre-structural through the extended abstract level in which students transfer from simple to complex applications. Raiyn and Tilchin (2016) propose a method for the adaptive complex assessment of HOTs through a problem base learning process. This is a three-stage assessment process that prompts the development of HOTs though each stage: 1) developing the HOTs, 2) developing the HOTs and collaborative skills, 3) assessment of the collaborative skills and construction of summative assessments of students (Raiyn & Tilchin, 2016). In addition to the PBL process which is student centered and adaptive in ways that allow for the development of students HOTs, researchers have also discussed the necessity of student engagement through choice and flexible assessments as methods through which students develop necessary HOTs.

Pretorius, van Mourik, & Barratt (2017) propose the development of flexible, student choice based assessment through which students are offered options and choose which to pursue. Biggs (2012) proposes that student engagement and buy in are considered central to effective educational practice. Authentic assessment task options presented to students allow them to see the transferability of skills being assessed to their future applications (Pretorius et al., 2017). When Pretorius et al. (2017) evaluated assessments based on both product-focused activities and process-focused activities, the assessment tools from the process focused (PBL type activities) were more effective in prompting deeper levels of (HOTs) thinking.

Through the careful examination of best practice in assessments and feedback regularly provided to students, educators can begin to address the gaps in practice of the

assessment of thinking. It is no longer an option to assess students using an unbalanced approach in which only LOTS are assessed using traditional standardized and summative measures. Educators must design learning opportunities that demonstrate students thinking capacity and their ability to apply what they have learning in a variety of setting and for a variety of purposes. These types of reflexive assessments and rubrics for the assessment of thinking can propel students to competencies in skills needed for the 21st Century and competition in a global economy (Jonsson, 2014).

Rubrics

The research on training teachers to use rubrics clearly demonstrates the need for comprehensive training in the use of rubrics to ensure the positive effects of rater reliability. Taylor and Galaczi (2011) discuss the need for comprehensive teacher training in rubrics based on the element of perception and the need to clarify evidence in student work when compared with rubric criteria. Often questioned in current research is how well teachers understand the constructs that are being assessed using a rubric and how this is an additional area in which teacher training is required when assessing students using criteria based rubrics (Yildiz, 2011). Bøhn (2018) holds that teachers as raters using rubrics effectively, can significantly impact student learning opportunities to establish genuine learning around HOTs.

The research is also clear on the importance of using rubrics. When comparing the benefits of rubrics to comprehensive graded category rating scales, Dogan and Uluman (2017) found that rubrics provide better access to consistent, genuine, formative assessment as a method of student feedback. Hassel (2015) found that measuring student

learning in a manner that provides clear criteria (a rubric) makes visible the measurement of student thinking and learning. While this more time consuming method of measurement is not without challenge, the tradeoff of effort is worthwhile as once the levels of expected proficiency have been delineated, genuine assessment of learning and progress can occur in an objective fashion.

The research is clear that rubrics can and should be used to measure HOTS. For example, Rembach and Dison (2016) studied the transformation of taxonomies into rubrics and demonstrated learning benefits in determining student's cognitive capacity when faced with set tasks. Constructive alignments (CA) between course descriptions, learning objectives, teaching and learning, and assessment must be interrelated for deep learning to occur (Biggs & Tang, 2011). Furthermore, Rembach and Dison (2016) note the promotion of HOTs when teachers, scorers, and students, had access to rubrics all the time to use as a feedback tool to gauge progress.

Recommendations for Further Study

The research completed and the assumptions drawn from the data collected and analyzed in my study provides a starting point for the continued development of teacher's awareness of HOTs and LOTs. Additionally, a critical component for future study is the implementation of continued professional development of teacher's capacity for providing learning experiences in which students can develop these HOTs and the methods by which they evaluate student success. While my study built awareness and interest in the field around the use of the BRT as a reliable and valid rubric for scoring student writing to determine if learning and thinking are occurring, it is imperative that

future studies continue around the necessary steps for providing the training that teachers need to implement the BRT as a rubric for scoring levels of thinking.

Knowing that building teacher awareness of HOTs and LOTs and the BRT as well as student's varied levels of thought will take sustained professional development opportunities, future research must continue around the most effective design to implement said training sessions. Once teachers have a firm understanding of the importance of HOTs and LOTs and the scaffolded implementations necessary for students to develop these skills, the focus must shift to the development of a BRT based rubric and sustained and evaluated professional development opportunities for teachers to practice implementation and use of the rubric to score writing for thinking.

Conclusions

The goal of this study was to fill a gap in educator practice for scoring student writing to include levels of student thinking using the BRT rubric. There is a risk when proposing a study that it is not grounded in familiarity or common practice, such as evaluating student thinking levels. The risk is that the study could be rejected by the participants. This risk was mitigated in two important ways. First, the development of higher order thinking skills is an espoused value of the involved schools. Second, the value is also an educational goal that is highly supported by the teachers who work at the schools. Teachers, however, would like to know that such evaluations are reliable and valid, and that desire reflects the purpose of this study. Teachers want to be able to evaluate how well their students are thinking. Teachers want to prepare students for success in the 21st century global economy. In an effort to continue this exploration to

determine a reliable and valid scoring tool for students thinking, it would be greatly appreciated if you would respond to the evaluation link accompanying the white paper in its original email body. The short, voluntary evaluation questions ask a) if you have any clarifying questions about the use of the BRT and the accompanying verb list as a scoring rubric for evaluating students level of thinking in writing, b) any suggestions you can make that would improve the BRT as a rubric for scoring student thinking through writing.

References

Aagaard, T., & Lund, A. (2013). Mind the gap: Divergent objects of assessment in technology-rich learning environments. *Nordic Journal of Digital Literacy, 8*(04), 225-243.

Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. Journal of The Medical Library Association, 103(3), 152-153. doi:10.3163/1536-5050.103.3.010

Argyris, C., & Schon, D. (1974). Theory in practice. San Francisco, CA: Jossey-Bass.

Al-Fayez, M. Q., & Jubran, S. M. (2012). The impact of using the heuristic teaching method on Jordanian mathematics students. *Journal of International Education Research*, 453-460. doi: http://dx.doi.org/10.19030/jier.v8i4.7293

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Boston, MA: Allyn & Bacon.

Atherton, J. S. (2013). Learning and teaching; SOLO taxonomy. Retrieved July 9, 2016 from http://www.learningandteaching.info/learning/solo.htm

Bartell, T. G. (2012). Learning to teach mathematics for social justice: Negotiating social justice and mathematical goals. National Council of Teachers of Mathematics, 44 (1), 129–163. Retrieved from www.nctm.org.

Benjamin, R., Miller, M. A., Rhodes, T. L., Banta, T. W., Pike, G. R., & Davies, G. (2012). *The seven red herrings about standardized assessments in higher education* (National Institute for Learning Outcomes Assessment, Occasional

Paper No.15). Retrieved from http://www.learningoutcomesassessment.org/occasionalpaperfifteen.htm

Bezuidenhout, M. J., & Alt, H. (2011). 'Assessment drives learning': Do assessments promote high-level cognitive processing? *South African Journal of Higher Education*, *25*(6).

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*, 347-364. doi: 10.1007/BF00138871

Biggs, J. B. & Tang, C. (2011). Teaching for quality learning at university. (4th Ed.). Maidenhead: McGraw Hill Education & Open University Press

Biggs, J. (2012). Enhancing learning through constructive alignment. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes.* (pp. 117–136). New York, NY: Cambridge University Press. https://doi-org.ezp.waldenulibrary.org/10.1017/CBO9781139048224.009

Bøhn, H. (2018). Assessing content in a curriculum-based EFL oral exam: The importance of higher-order thinking skills. *Journal Of Language Teaching & Research*, *9*(1), 16-26. doi:10.17507/jltr.0901.03

Brent, D. (2011). Transfer, transformation, and rhetorical knowledge: Insights from transfer theory. *Journal of Business and Technical Communication, 25*, 396-420. doi: 10.1177/1050651911410951

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Alexandria, VA: ASCD.

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review, 67*(3), 343-368.

Brown, C. (2016). *Patterns of innovation: Showcasing the nation's best in 21st centurylLearning*. Washington, DC: Pearson Foundation.

Choudhury, B., Gouldsborough, I., & Shaw, F. L. (2015). The intelligent anatomy spotter: A new approach to incorporate higher levels of Bloom's taxonomy. *Anatomical Sciences Education, 9*, 440-445. doi: 10.1002/ase.1588

Comparative & International Educational Society. (2014). *Renvisioning education for all.* Toronto, Canada.

Clark, I. (2011). Formative assessment: Policy, perspective, practice. *Florida Journal of Education Administration and Policy, 4*(2), 158-180.

Collet, V. S. (2014). The GIR model: Mentoring for teacher effectiveness. *English Leadership Quarterly, 37*(2), 9-13. Retrieved from http://www.ncte.org/journals/elq/issues/

Common Core State Standards Initiative. (2010). *Common core state standards for mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Conley, D. (2015). A new era for educational assessment. *Education Policy Analysis Archives, 23*(8), 1-36. doi: http://dx.doi.org/10.14507/epaa.v23.1983.

Crockett, L. (2017) Bloom's Digital Taxonomy Verbs. *Global Citizenship.* Retrieved from https://wabisabilearning.com/resources/blooms-digital-taxonomy-verbs

Cunningham, J., & De Aquino, C. E. (2015). A brief history of competency-based learning in the united states. *Journal On Management, 10*(3), 1-5.

Dagostino, L., Carifio, J., Bauer, J. D., Zhao, Q., & Hashim, N. H. (2015). Using Bloom's Revised Taxonomy to analyze a reading comprehension instrument. *Current Issues in Education, 18*(2). Retrieved from http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1379

Dogan, C. D., & Uluman, M. (2017). A comparison of rubrics and graded category rating scales with various methods regarding raters' reliability. *Educational Sciences: Theory And Practice*, *17*(2), 631-651.

Dolan, E. L., & Collins, J. P. (2015). We must teach more effectively: Here are four ways to get started. *Molecular Biology of the Cell*, *26*, 2151-2155. doi: 10.1091/mbc.E13-11-0675

Draper, D. C. (2015). Collaborative instructional strategies to enhance knowledge convergence. *American Journal of Distance Education*, *29*, 109-125. doi: http://dx.doi.org/10.1080/08923647.2015.1023610

Fink, D. L. (2003) *Creating significant learning experiences: An integrated approach to designing*. Hoboken, NJ: Jossey-Bass.

Forehand, M. (2010). Bloom’s taxonomy. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology* (pp 41-47). Retrieved from http://www.palieducationsociety.org/images/ebooks%20(13).pdf

Gallavan, N. P., & Kottler, E. (2012). Advancing social studies learning for the 21st century with divergent thinking. *The Social Studies*, 165-170. doi: http://dx.doi.org/10.1080/00377996.2011.605641

Gardner, H. (2010). Five minds for the future. In J. Bellanca and R. Brandt (Eds.), *21st Century Skills: Rethinking How Students Learn* (pp. 9-32). Bloomington, IN: Solution Tree Press.

Ganapathy, M., & Wai Kit, L. (2017). Promoting HOTS via ICT in ESL classrooms. *The Seventh International Language Learning Conference Report.*

Gerard, L. l., Kihyun, R., McElhaney, K. W., Liu, O. L., Rafferty, A. N., & Linn, M. C. (2016). Automated Guidance for Student Inquiry. *Journal Of Educational Psychology*, *108*(1), 60-81.

Günel, M., Memis, E. K., & Büyükkasap, E. (2010). Effects of the science writing heuristic approach on primary school students' science achievement and attitude toward science course. *Egitim ve Bilim*, *35*(155), 49-62. Retrieved from https://www.researchgate.net

Halpern, D. F. (2013). *Thought and knowledge: An introduction to critical thinking*. Psychology Press.

Haolader, F. A., Avi, M. R., & Foysol, K. M. (2015). The taxonomy for learning, teaching and assessing: Current practices at polytechnics in Bangladesh and its effects in developing students' competences. *International Journal for Research in Vocational Education and Training*, 99-118. doi: http://dx.doi.org/10.13152/IJRVET.2.2.9

Harvey, S., & Daniels, H. (2009). *Comprehension & collaboration: Inquiry circles in action*. Portsmouth, NH: Heinemann.

Hassel, H. (2015). Analyzing evidence with rubrics. *Teaching English in the Two Year College, 43*(2), 202-205.

Haynes, A., Lisic, E., Goltz, M., Stein, B., & Harris, K. (2016). Moving Beyond Assessment to Improving Students' Critical Thinking Skills: A Model for Implementing Change. *Journal of The Scholarship Of Teaching & Learning*, *16*(4), 44-61.

Heick, T. (2016). *Teach thought*. Retrieved October 7, 2016 from http://www.teachthought.com/category/critical-thinking/blooms-taxonomy/.

Herman, J., Linn, R., & Moss, F. (2013). *On the road to assessing deeper learning: The status of smarter balanced and PARCC assessment consortia* (CRESST Report 823). Retrieved from http://cresst.org/wp-content/uploads/R823.pdf

Hess, K. K., Jones, B. S., Carlock, D., & Walkup, J. R. (2009). *Cognitive rigor: Blending the strengths of Bloom's taxonomy and Webb's depth of knowledge to enhance classroom-level processes*. Retrieved from ERIC database.

Hong, Y. C., & Choi, I. (2011). Three dimensions of reflective thinking in solving design problems: A conceptual model. *Educational Technology Research and Development*, 687-710. doi: 10.1007/s11423-011-9202-9

Huitt, W. (2011). *A holistic view of education and schooling: Guiding students to develop capacities, acquire virtues, and provide service*. Retrieved from http://www.edpsycinteractive.org/papers/holistic-view-of-schooling-rev.pdf

Hung, W. (2006). The 3C3R model: A conceptual framework for designing problems in PBL. *Interdisciplinary Journal of Problem-based Learning, 1*(1). doi: http://dx.doi.org/10.7771/1541-5015.1006

Hurson, T. (2008). *Think better: An innovator's guide to productive thinking*. New York, New York: McGraw-Hill.

Jailani, & Retnawati, H. (2016). The challenges of junior high school mathematic teachers in implementing the problem-based learning for improving the higher-order thinking skills. *The Online Journal of Counseling and Education, 5*(3), 1–13.

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation In Higher Education*, *39*(7), 840-852.

Kaiser, L., Kaminski, K., & Foley, J. (2013). *Learning transfer in adult education: New directions for adult and continuing education*. Danvers, MA: Jossey Bass.

Kaldor, E. (2018) *Higher and lower order thinking skills.* Supporting Teaching Excellence at URI.

Kaufman, J. C., Lee, J., Baer, J., & Lee, S. (2007). Captions, consistency, creativity, and the consensual assessment technique: New evidence of reliability. *Thinking Skills And Creativity, 2*(2), 96-106.

Kolb, D. (1984). Experiential learning as the science of learning and development.

Lam, R. (2017). 'Taking stock of portfolio assessment scholarship: From research to practice'. *Assessing Writing, 31*, 84–97.

Lo, S. M., Larsen, V. M., & Yee, A. T. (2016). A two-dimensional and non-hierarchical framework of Bloom's taxonomy for biology [Abstract]. *The FASEB Journal, 30*(1 Supplement), 662.14. Retrieved from http://www.fasebj.org/content/30/1_Supplement/662.14

McMillan, J. H. (2013). *Classroom assessment: Pearson new international edition: Principles and practice for effective standards-based instruction* [Kindle version]. Retrieved from Amazon.com

McNamara, C. (2006). *Field guide to consulting and organizational developmental: A collaborative and systems approach to performance, change and learning.* Authenticity Consulting.

Mokharti, K., Yellin, D., Bull, K., & Montgomery, D. (1996). Portfolio assessment in teacher education: Impact on preservice teacher's knowledge and attitudes. *Journal of Teacher Education*, 47 (4), 245-252.

Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. Retrieved from https://ebookcentral.proquest.com

Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, 1-36. doi: http://dx.doi.org/10.1080/13546780802490186

Nkhoma, M. Z., Lam, T. K., Sriratanaviriyakul, N., Richardson, J., Kam, B., & Lau, K. H. (2017). Unpacking the revised Bloom's taxonomy: developing case-based learning activities. *Education + Training*, *59*(3), 250-264. doi:10.1108/ET-03-2016-0061

Panasan, M., & Nuangchalerm, P. (2010). Learning outcomes of project-based and inquiry-based learning activities. *Journal of Social Sciences, 6*(2), 252-255.

Peen, T. Y., & Arshad, M. Y. (2014). Teacher and student questions: A case study in Malaysian secondary school problem-based learning. *Asian Social Science, 10*(4), 174. doi: http://dx.doi.org/10.5539/ass.v10n4p174

Pretorius, L., van Mourik, G. P., & Barratt, C. (2017). Student choice and higher-order thinking: Using a novel flexible assessment regime combined with critical thinking activities to encourage the development of higher order thinking. *International Journal Of Teaching & Learning In Higher Education, 29*(2), 389-401.

Purnomo, Y. W. (2017). The complex relationship between teachers' mathematics-related beliefs and their practices in mathematics class. *New Educational Review, 47*(1), 200–210. doi:10.15804/ tner.2017.47.1.16.

Raiyn, J., & Tilchin, O. (2016). The impact of adaptive complex assessment on the HOT skill development of students. *World Journal Of Education*, *6*(2), 12-19.

Rembach, L., & Dison, L. (2016). Transforming taxonomies into rubrics: Using SOLO in social science and inclusive education. *Perspectives in Education*, *34*(1), 68-83.

Retna, K. S., & Ng, P. T. (2016). The application of learning organization to enhance learning in Singapore schools. *Management in Education*, *30*(1), 10-18. doi: 10.1177/0892020615619665.

Retnawati, H. H., Djidu, H. H., Kartianoml, K., Apino, E. A., & Anazifa, R. R. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems Of Education In The 21St Century*, *76*(2), 215-230.

Rothstein, D., & Santana, L. (2011). *Make just one change: Teach students to ask their own questions*. Cambridge, MA: Harvard Education Press.

Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice* (2nd ed.). Amsterdam: Academic Press, an imprint of Elsevier.

Scott, L. A. (2017). 21st century skills early learning framework. Partnership for 21st Century Skill (P21). Retrieved from http://www.p21.org/storage/documents/EarlyLearning_Framework/ P21_ELF_ Framework_Final.pdf.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29, 4-14.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., Martinez, J. L., & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts, 2*, 68-85. doi: http://dx.doi.org/10.1037/1931-3896.2.2.68

Sotiriadou, P., & Hill, B. (2015). Using scaffolding to promote sport management graduates' critical thinking. *Annals of Leisure Research*, 105-122. doi: http://dx.doi.org/10.1080/11745398.2014.925406

Spruce, R., & Bol, L. (2015). Teacher beliefs, knowledge, and practice of self-regulated learning. Metacognition And Learning, 10(2), 245-277.

Sultana, F. (2010). *An initial study of a method for instructing educators about the revised taxonomy*. Retrieved from ERIC database. (ED523761)

Taylor, L., & Galaczi, E. (2011). *Scoring validity. Studies in language testing. Examining speaking. Research and practice in assessing second language speaking*, Cambridge: Cambridge University Press.

Tidwell, A. L. (2015). Assessing student problem solving using structured versus unstructured case analysis. *Academy of Educational Leadership Journal*, *19*(3), 309-324.

Tíjaro-Rojas, R., Arce-Trigatti, A., Cupp, J., Pascal, J., & Arce, P. E. (2016). A systematic and integrative sequence approach (SISA) for mastery learning: Anchoring Bloom's Revised Taxonomy to student learning. *Education for Chemical Engineers*, *17*, 31-43. doi: http://dx.doi.org/10.1016/j.ece.2016.06.001

Thompson, R., & O' Loughlin, D. (2015). The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anatomical Sciences Education*, *8*, 493-501. doi: 10.1002/ase.1507

Vista, A., Care, E., & Griffin, P. (2015). A new approach towards marking large-scale complex assessments: Developing a distributed marking system that uses an automatically scaffolding and rubric-targeted interface for guided peer-review. *Assessing Writing*, *24*, 1-15. doi: http://dx.doi.org/10.1016/j.asw.2014.11.001

Wagner, T. (2014). *The global achievement gap* (Upd. Rev. Ed.). New York, NY: Basic Books.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Yildiz, L. M. (2011). English VG1 level oral examinations: How are they designed, conducted and assessed? (Unpublished). University of Oslo, Oslo, Norway.

Zull, J. E. (2006). Key aspects of how the brain learns. *New directions for Adult and Continuing Education, 2006*(110), 3-9. doi: 10.1002/ace.213

Zull, J. E. (2011). *From brain to mind: Using neuroscience to guide change in education*. Sterling, VA: Stylus Publishing, LLC.

Appendix B: Levels of Bloom's Revised Taxonomy

	Levels of Bloom's Revised Taxonomy
6	Creating
5	Evaluating
4	Analyzing
3	Applying
2	Understanding
1	Remembering

Note. From "Bloom's Revised Taxonomy (BRT) Verbs and Similar Verbs from Higher Order Thinking Skills (HOTS) down to Lower Order Thinking Skills (LOTS)," by Crockett, *Global Citizen*, 2017, https://globaldigitalcitizen.org/category/blooms-taxonomy

Appendix C: Bloom's Revised Taxonomy (249 Action Verbs)

Knowledge	Understand	Apply	Analyze	Evaluate	Create
define	explain	solve	compare	reframe	design
identify	describe	apply	classify	criticize	compose
describe	interpret	illustrate	contrast	evaluate	create
label	paraphrase	modify	distinguish	order	plan
list	summarize	use	infer	appraise	combine
name	classify	calculate	separate	judge	formulate
state	compare	change	explain	support	invent
match	differentiate	choose	select	compare	hypothesize
recognize	discuss	demonstrate	categorize	decide	sustitute
select	distinguish	discover	connect	discriminate	write
examine	extend	experiment	differentiate	recommend	compile
locate	predict	relate	discriminate	summarize	construct
memorize	associate	show	divide	assess	develop
quote	contrast	sketch	order	choose	generalize
recall	convert	complete	point out	convince	integrate
reproduce	demonstrate	construct	prioritize	defend	modify
tabulate	estimate	dramatize	subdivide	estimate	organize
tell	express	interpret	survey	find errors	prepare
copy	identify	manipulate	advertise	grade	produce
discover	indicate	paint	appraise	measure	rearrange
duplicate	infer	prepare	break down	predict	rewrite
enumerate	relate	produce	calculate	rank	role-play
listen	restate	report	conclude	score	adapt
observe	select	teach	correlate	select	anticipate
omit	translate	act	criticize	test	arrange
read	ask	administer	deduce	argue	assemble
recite	cite	articulate	devise	conclude	choose
record	discover	chart	diagram	consider	collaborate
repeat	generalize	collect	dissect	critique	collect
retell	give examples	compute	estimate	debate	devise
visualize	group	determine	evaluate	distinguish	express
	illustrate	develop	experiment	editorialize	facilitate
	judge	employ	focus	justify	imagine
	observe	establish	illustrate	persuade	infer
	order	examine	organize	rate	intervene
	report	explain	outline	weigh	justify
	represent	interview	plan		make
	research	judge	question		manage
	review	list	test		negotiate
	rewrite	operate			originate
	show	practice			propose
	trace	predict			reorganize
	transform	record			report
		schedule			revise
		simulate			schematize
		transfer			simulate
		write			solve
					speculate
					structure
					support
					test
					validate

Note. From "249 Bloom's Taxonomy Verbs for Critical Thinking," by Teach Thought Staff, 2017 https://www.teachthought.com/critical-thinking/249-blooms-taxonomy-verbs-for-critical-thinking/

Appendix D: Sample E-mail Contact to Potential Participants

Dear Potential Participant,

I hope you are having a lovely academic school year! I have been attempting to get outside and enjoy the beautiful weather while simultaneously working on my doctoral research study. As a fellow educator, I truly value your commitment and dedication to the field and also want to honor your time and commitments. If you feel it will not be a burden, I am asking that you read through my research consent form in this email. If after reading through the information related to my research study and you would like to volunteer to participate in the study please follow the directions to give consent.

The study has been approved by Walden University.

Thank you very much for your time and consideration!

Sincerely,

Siri DeForest Reynolds

Appendix E: Participant Agenda

10:00am Check-in
10:05am Review agenda
10:10am Begin Training Session
10:15am Overview of Bloom's Revised Taxonomy (BRT)
-listing key components/levels
-brief review of Bloom's Taxonomy and the transition to BRT

10:20am Overview of BRT Scoring Sheet & the sheet of 249 verbs similar to the BRT levels

10:25am Sorting exercise 1
Exercise instructions: *"Please sort and arrange in order the sentence strips in envelope #1 according the level of BRT you feel they best align with".*
10:35am Sorting exercise 2
Exercise instructions: *"Please sort and arrange in order the sentence strips in envelope #2 according the level of BRT you feel they best align with".*
10:45am Scoring warm up 1
Exercise instructions:
"Read over the BRT, look at the verbs. Next, read the whole text, then re-read and integrate BRT levels, for further clarification read through the BRT 249 verbs, you can be flexible, this is the suggestion for scoring.

There is no one right way to use these tools for scoring writing, you should feel free to mark up your hard copies while processing which level to input into the GoogleSheet as a BRT level score for that writing sample.

Please use the BRT Scoring Sheet to score the writing sample in envelope #3".
10:55am Questions session regarding procedures for coding student writing samples
11:00am Scoring session begins
Once the coding session begins I am unable to answer questions regarding the coding process or any of the student writing samples. From this point forward I will ensure that there is not talking or communication between participants, will remind them of break and refreshment times throughout the coding session.
Breaks will occur every hour and will include snacks and drinks as well as a lunch break during which sandwiches, fruit and drinks will be served.
11:00-3:00 Scoring of writing samples

Upon the completion of the scoring of each writing sample, each participant will receive their $20 Starbucks gift card and will be dismissed.

Appendix F: Research and Project Progress

Data Collection Training Reflection

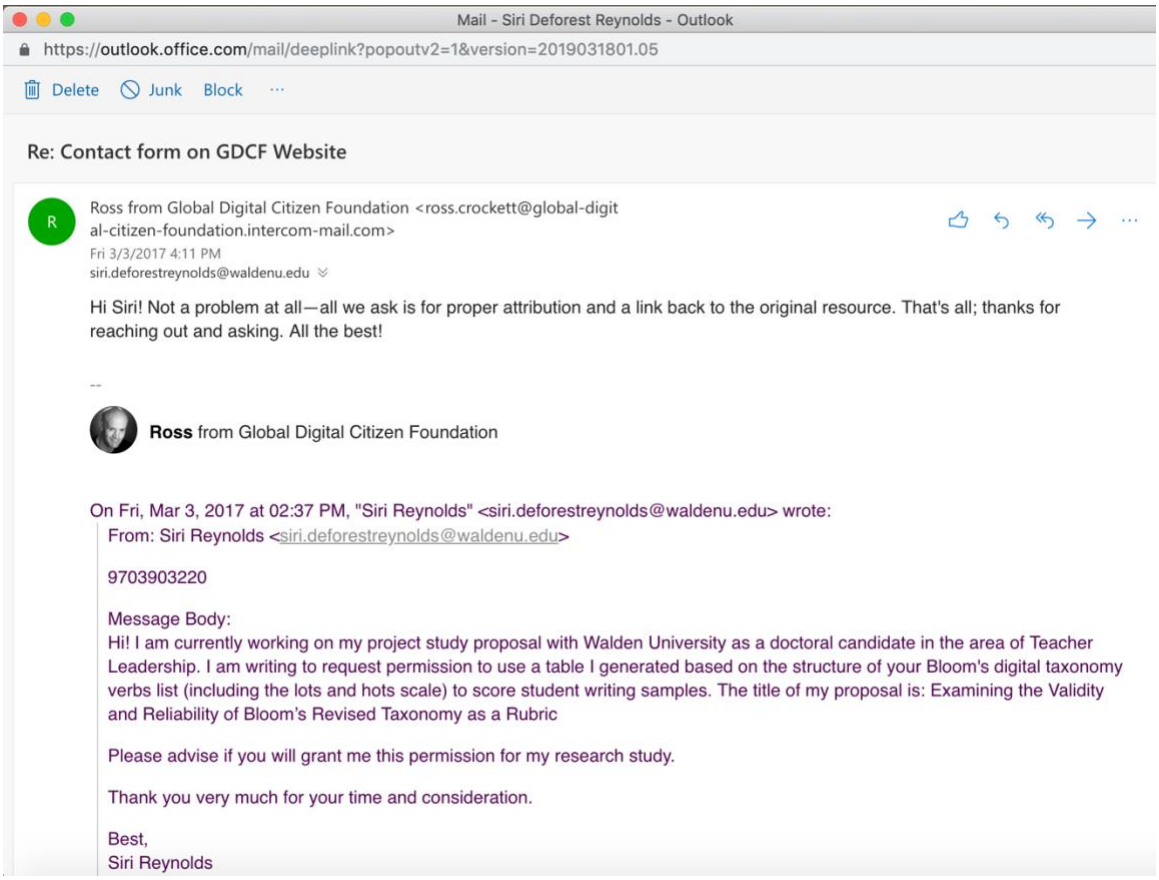
Upon gathering in the conference room and commencing the session, I reviewed the agenda. The tone of the group was inquisitive and prepared to engage. The study participants did not ask questions during the review of the agenda. After reviewing the agenda I passed out a sample of the BRT scoring sheet and the 249 verbs associated with the BRT. I shared with the group the process through which Krathwol initiated the revision of the BRT to include the create level of thinking. A participant asked at this time asked if there were any known methods for scoring student writing for levels of thinking at which I reiterated the purpose of the study to help contribute to the field in this manner. Speaking directly to the purpose of examining student thinking and how it develops over time rather than focusing on the regurgitation of the content as a method of evaluating student learning and academic capacity. After this discussion and the group spent about 10 minutes reviewing the BRT overview as well the associated verbs. Once the group determined they did not need any further clarification, we began the first training exercise.

The group began sorting the sentence strips according to the BRT levels to which they most closely associated. They generally determined the same levels without speaking which seemed promising. After I determined the group had completed the first training exercise I moved the group on the second training exercise. The participants associated these sentence strips less closely to each others results. Hindsight shows the writing samples likely influenced the ability of the participants to associate sentences

with the levels of the BRT. At this time a participant asked why the verbs in the BRT levels were repeated. The discussion between the participants drew attention to the application of the verb in each level. For example the verb, evaluate was listed in three categories and in each the participants identified a manner in which evaluation would be appropriate.

After completing the second exercise, the group moved to the scoring the writing sample section of the training. A participant asked at this time if they were looking for the highest possible level of the BRT attained in the writing of the sample. Using the BRT verbs each participant spent time looking back and forth between the verbs and the writing sample. At this time, I reiterated the fact there is no correct way of scoring these writing samples using the BRT and that this study will provide insight into the examination for the reliability and validity of using the BRT as a scoring tool.

Appendix G: Permission Letter from Global Citizen



Mail - Siri Deforest Reynolds - Outlook

https://outlook.office.com/mail/deeplink?popoutv2=1&version=2019031801.05

Delete Junk Block ···

Re: Contact form on GDCF Website

R

Ross from Global Digital Citizen Foundation <ross.crockett@global-digital-citizen-foundation.intercom-mail.com>

Fri 3/3/2017 4:11 PM

siri.deforestreynolds@waldenu.edu

Hi Siri! Not a problem at all—all we ask is for proper attribution and a link back to the original resource. That's all; thanks for reaching out and asking. All the best!

--

Ross from Global Digital Citizen Foundation

On Fri, Mar 3, 2017 at 02:37 PM, "Siri Reynolds" <siri.deforestreynolds@waldenu.edu> wrote:

From: Siri Reynolds <siri.deforestreynolds@waldenu.edu>

9703903220

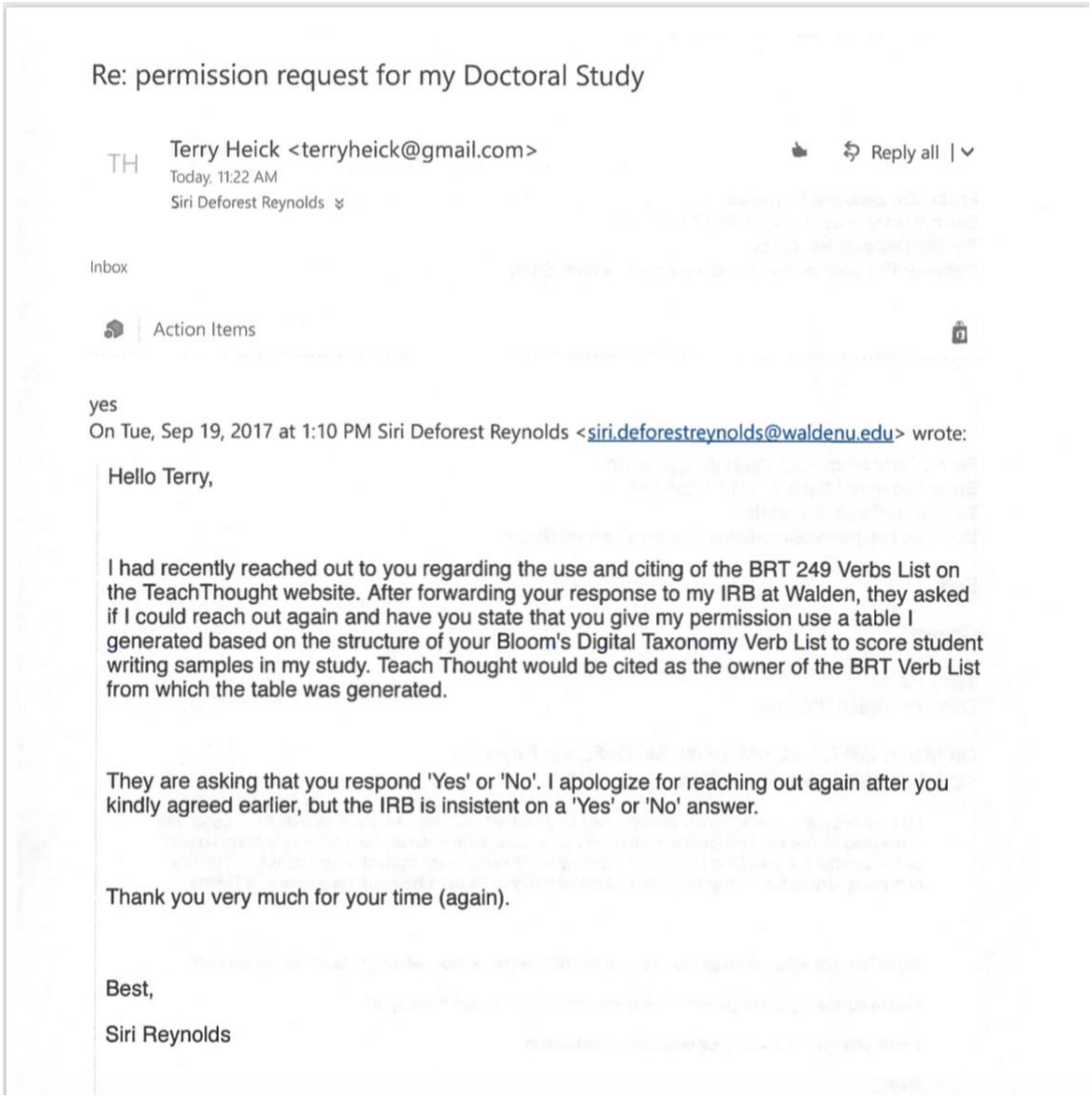
Message Body:
Hi! I am currently working on my project study proposal with Walden University as a doctoral candidate in the area of Teacher Leadership. I am writing to request permission to use a table I generated based on the structure of your Bloom's digital taxonomy verbs list (including the lots and hots scale) to score student writing samples. The title of my proposal is: Examining the Validity and Reliability of Bloom's Revised Taxonomy as a Rubric

Please advise if you will grant me this permission for my research study.

Thank you very much for your time and consideration.

Best,
Siri Reynolds

Appendix H: Permission Letter from Teach Thought



Re: permission request for my Doctoral Study

TH Terry Heick <terryheick@gmail.com>
Today, 11:22 AM
Siri Deforest Reynolds

Reply all

Inbox

Action Items

yes
On Tue, Sep 19, 2017 at 1:10 PM Siri Deforest Reynolds <siri.deforestreynolds@waldenu.edu> wrote:

Hello Terry,

I had recently reached out to you regarding the use and citing of the BRT 249 Verbs List on the TeachThought website. After forwarding your response to my IRB at Walden, they asked if I could reach out again and have you state that you give my permission use a table I generated based on the structure of your Bloom's Digital Taxonomy Verb List to score student writing samples in my study. Teach Thought would be cited as the owner of the BRT Verb List from which the table was generated.

They are asking that you respond 'Yes' or 'No'. I apologize for reaching out again after you kindly agreed earlier, but the IRB is insistent on a 'Yes' or 'No' answer.

Thank you very much for your time (again).

Best,

Siri Reynolds