



Examining Score Report Language in Accountability Testing

Mary Roduta Roberts^{1*}, Chad M. Gotch² and Jessica Nina Lester³

¹ Department of Occupational Therapy, University of Alberta, Edmonton, AB, Canada, ² Learning and Performance Research Center, Washington State University, Pullman, WA, United States, ³ Department of Counseling and Educational Psychology, University of Indiana Bloomington, Bloomington, IN, United States

The purpose of this study was to examine patterns of language use in score reports within a North American context. Using a discourse analysis approach informed by conversation analysis, we explored how language was structured to express ideas or (re-)produce values, practices, and institutions in society. A sample of 10 reports from the United States and Canada within the domain of accountability testing was selected. Observed patterns of language represented micro-discourses embedded within broader discourse related to accountability mandates within each country. Three broad themes were identified within and across the score reports—Displays of Information, Knowledge Claims, and Doing Accountability. Within each of the broad themes were sub-themes—word choice, and visual representations, script formulations, hedging, establishing authority, and establishing responsibility—that characterized more fine-grained textual features. Future research may explore empirical evidence for the social dynamics identified through this study's textual analysis. Complementary lines of research on cognitive, affective, and socio-cultural factors of score report interpretation and use are encouraged.

OPEN ACCESS

Edited by:

Mary Frances Hill,
University of Auckland, New Zealand

Reviewed by:

Robin Dee Tierney,
Research-for-Learning, United States
Catarina Andersson,
Umeå University, Sweden

*Correspondence:

Mary Roduta Roberts
mroberts@ualberta.ca

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 30 December 2017

Accepted: 23 May 2018

Published: 19 June 2018

Citation:

Roduta Roberts M, Gotch CM and
Lester JN (2018) Examining Score
Report Language in Accountability
Testing. *Front. Educ.* 3:42.
doi: 10.3389/feduc.2018.00042

Keywords: score report, discourse analysis, conversation analysis, activity theory, accountability

INTRODUCTION

Over the past two decades there has been a movement in North American educational systems to emphasize accountability roles within student assessment (McCall, 2016; Wise, 2016). The most notable push for this emphasis was found in the No Child Left Behind (NCLB) Act in the United States (United States Congress, 2002). This act mandated for every state in the union comprehensive standardized testing from grades 3 through 8 and once again in high school with stakes attached to poor performance, such as turnover of school administration. In some states additional stakes, such as retention of the student at grade 3, termination of the teachers of poor-performing students, and denial of the high school diploma were added on top of those specified in the federal statute (National Conference of State Legislatures, 2017; Washington State Legislature, 2017). Comprehensive, accountability-focused components of assessment systems were also taken up in Canadian provinces such as Alberta and Ontario (Legislative Assembly of Ontario, 1996; Legislative Assembly of Alberta, 2017). While differences may exist in the specific mechanisms

and mandates of accountability and, more broadly, orientations toward high-stakes testing in the United States and Canada, at a broad level the two countries share a trajectory toward increased oversight of the use of public funding for education. Within this push, governmental education agencies have leveraged large-scale testing as a tool to hold educational systems accountable for student learning against stated benchmarks.

Highly visible accountability tests have introduced to the educational experience annual (and often more frequent) communication to the student that identifies their status against a standard of proficiency. Such communication has often taken the form of a physical score report that is delivered to the student and their family by the state/provincial education agency. In many cases, a contracted testing company may collaborate with the state/provincial education agency or lead development of the report. Common report content includes a description of the student's test performance in numeric, graphic, and text-based forms. Questions persist regarding how well stakeholders understand the purpose of the tests and the content of the communications (i.e., score reports) stemming from these testing programs (e.g., Hambleton and Slater, 1997; Jones and Desbiens, 2009; Alonzo et al., 2014). Several studies have examined how stakeholders interpret the information presented in score reports (e.g., Whittaker et al., 2011; van der Kleij and Eggen, 2013; Zwick et al., 2014). Score report *use* has been examined less frequently (Gotch and Roduta Roberts, in press), but is still held, either implicitly or explicitly, as an ultimate aim of report development. For example, Zenisky and Hambleton (2015) note guidance on test score use, next steps, and links to external resources as potential descriptive elements of score reports. With residence in systems of accountability but adoption of a mission to be used to further student learning, score reports can occupy both summative and formative roles. In this paper, we consider how the written content of score reports may function within the accountability sphere of contemporary educational systems in North America, specifically the Anglophonic regions of the United States and Canada.

Communicating Results From a Psychometric Perspective

Score reports have traditionally been framed using a psychometric perspective, residing within the larger process of test development (e.g., Hambleton and Zenisky, 2013). Essential to this perspective has been the role of score reporting to the validity of its testing program (O'Leary et al., 2017). Such importance is reflected in the *Standards for Educational and Psychological Testing*, which contain several entries that address score reporting directly (e.g., 6.0, 6.10–6.13; American Educational Research Association et al., 2014). Accordingly, most advances to-date have focused primarily on methods for presenting information about test performance in ways that maximize accurate interpretation (e.g., dynamic reporting elements, Zapata-Rivera and VanWinkle, 2010; online tutorials, Zapata-Rivera et al., 2015). In line with this validity-oriented emphasis, current approaches to the communication of assessment results primarily focus on the transmission and

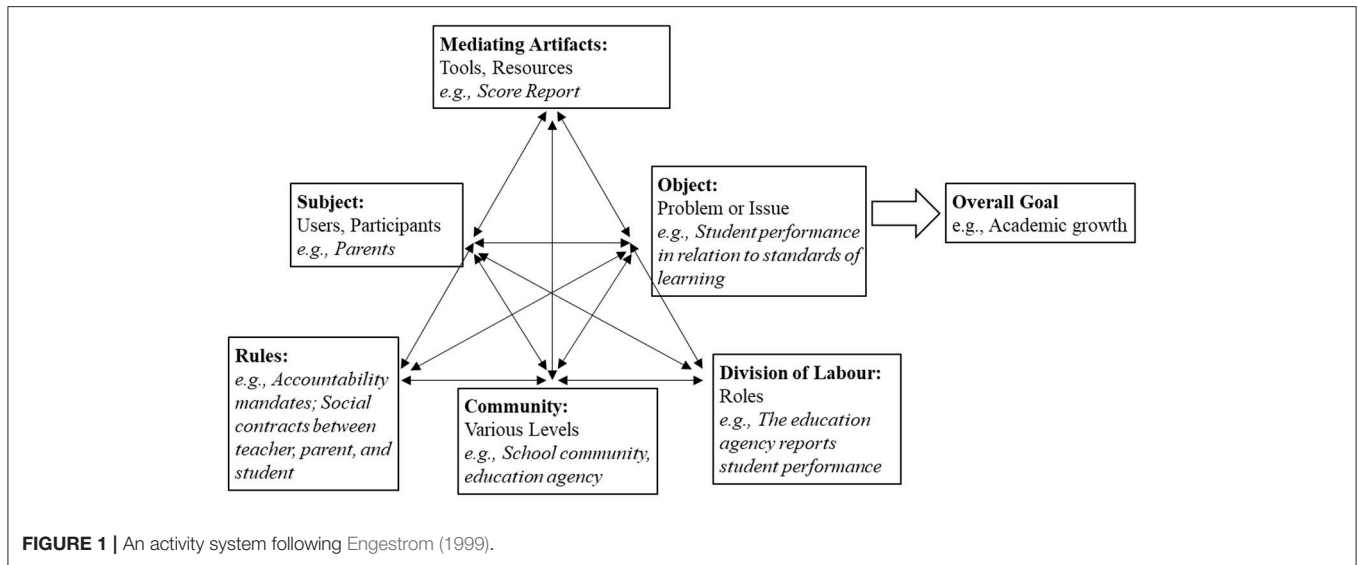
processing of information by the end users (Behrens et al., 2013; Gotch and Roduta Roberts, 2018). In this frame of communication, information flows from the test developer to the user, and an effective transmission has taken place when the information is understood by the user in accordance with the intentions of the test developer.

While the information-processing perspective is useful for examining interpretations formed by an individual in response to being presented with a score report, it cannot directly address issues of score report use. Given that use is a form of activity stakeholders engage in collectively when interacting with an assessment system, a thorough consideration of use may require application of a more socially oriented frame. By shifting from the information-processing/psychometric perspective to an activity theory perspective (Engestrom, 1999; Behrens et al., 2013) we broaden the lens from consideration of the individual to the larger context or system within which the individual operates.

From the Individual to Social: An Alternative Frame of Reference for Communication of Assessment Results

An educational system comprises multiple actors that interact to direct, implement, and evaluate the delivery of educational programs. Embedded within this system are the policymakers, assessment developers, educators, and students, each operating within the shared context of the educational system but also within their own local contexts (e.g., governance, running a testing program, teaching, and learning within a classroom). Taking the view of this system, communication of assessment results can be recast from the output of a development process from a particular testing program (i.e., a local context) to playing an important supporting role within the larger educational system. For example, score reports could function in a role of assisting stakeholders, such as students, achieve mastery of the learning standards. Activity theory (Engestrom, 1999, 2006) provides a framework for examining the collective use of score reports to achieve this goal. Activity theory invokes sociological and anthropological concepts espousing a view where all activity is intentional, goal-oriented, mediated, and embedded within a particular context. To understand activity, one must consider multiple dimensions which are conceptually represented within an *activity system*.

Figure 1 depicts an activity system oriented around characterizing student performance in relation to standards of learning with an overall goal of academic growth for the student. The top half of the triangle in the activity system diagram illustrates the interaction between a person or group of participants (Subject), a problem or issue (Object), and tools or resources (Mediating Artifacts) when engaging in an activity. Subjects act on meaningful problems or issues usually through tools, which can be cognitive or physical, conceptual, or tangible. Activity evolves over time in a developmental manner with incremental changes within the system (e.g., by virtue of completing the goal; Engestrom, 1999). The activity system also captures the larger context within which an activity occurs, as represented by elements in the bottom half of the



triangle, including social rules or norms, the larger community to which the subject belongs, and a division of labor among those involved. Meaningful activity may involve multiple people in its production, and what constitutes as meaningful is bound by the social context (Jonassen and Rohrer-Murphy, 1999; Engeström, 2006). Of note in the diagram are the reciprocal relationships that exist between all points in the activity system, representing the various mutual influences upon each other.

Score Reports and Communication Theory

Positioning score reports as mediating artifacts within an educational system necessitates a view of communication that is different from the information-processing perspective typically adopted. Numerous traditions exist within communication theory that describe different purposes and functions of communication (Craig, 1999). The semiotics and socio-cultural traditions of communication align well with a conceptualization of score reporting within activity theory. The semiotics tradition is characterized by the use of signs and symbols to express ideas, achieve shared meaning, or to connect with others. The socio-cultural tradition is characterized by the processes that produce and reproduce the values, practices, institutions, and other structures within society. Both traditions are relevant in the characterization of score reporting as communication. Semiotics invites us to look closer at how common conventions for reporting assessment results, such as graphs or charts, function to meaningfully connect with various educational stakeholders. A socio-cultural perspective assists an examination of how current practice may result in the development of score reports that reflect or reinforce the values and expectations of the system which created them. The introduction of alternative lenses to frame communication provides us with multiple perspectives by which to view score reporting as an activity, including its development and outcomes.

When score reports are positioned as mediating artifacts within an educational activity system, the application of the

semiotics and socio-cultural perspectives of communication invites us to examine the core of the report itself—language. Examination of language used within the score report includes the text used to summarize student performance, the graphical displays used to communicate information in different ways, and an understanding that these language choices reflect the culture and norms of the system that produced the report. Activity theory, as an overarching framework, sets the context for application of communication perspectives to score reporting. In particular, activity theory provides insight into the context of language production. Rather than a focus on interpretability of language, analysis using this framework provides a window into how different individuals (i.e., subjects) who are all working toward the same overall goal within a system may have different priorities and motives, perhaps acting on different objects within the system. Activity theory-driven analysis may also reveal rules and roles defined by the community that can influence actions on the object. In summary, application of communication theories and activity theory provides a means for understanding the social forces that mediate score report development and use.

Study Purpose

This study was part of a larger effort to examine score reporting within the genre of accountability testing in K-12 education in Canada and the United States. The purpose of the present study was to examine patterns of language use in score reports within this North American context. Specifically, we explored how language was structured to express ideas or (re-)produce values, practices, and institutions in society. The former focus aligns with the semiotic tradition of communication; the latter aligns with the socio-cultural tradition. Throughout, score reporting is framed as part of an educational activity system including, for example, students as subjects, implied rules for student engagement as defined by the larger educational community that includes teachers and policymakers, and division of labor or community member roles and responsibilities in the achievement

of educational outcomes. We explicitly note this study was intended to be descriptive, not critical or evaluative of score reporting practices. Given our reading of the score reporting research to date and noting issues that have remained similar over time, an examination of language use through application of an alternative frame of reference has potential for illuminating reasons of why such issues have persisted. These reasons emerge in closer examination of the activity system and identification of areas of tension motivating future studies on score report development, interpretation, and use.

METHODS

Data Collection

English-language individual student score reports ($N = 40$) for elementary-level accountability tests representing 42 states in the United States and 5 provinces in Canada were obtained in electronic form from state and provincial department of education websites. This sample included all readily available sample reports, and served a line of research centered on the assessment of score report quality through a traditional psychometric lens (Gotch and Roduta Roberts, 2014; Roduta Roberts and Gotch, 2016). From the total pool of 40 reports, a sample of 10 reports was selected for the present study. We employed a purposeful sampling approach to select these reports. Notably, the criteria used to select the reports centered on our analytic aim of analyzing the language practices. Thus, we selected reports that were crafted in such a way that we could analyze the language choices. Specifically, we only considered those score reports that included ample text-based content in the form of an introductory letter and/or description of the testing program, score interpretations, or descriptions of performance levels. No consideration was given to the quality of the content, nor were any a priori expectations of language themes applied to selection of the report. In carrying out this process, the second author identified candidate reports and shared them with the other two authors. A consensus was reached to study 10 reports obtained from Arkansas, Delaware, Hawaii, Iowa, Maryland, Massachusetts, Missouri, Montana, North Dakota, and Ontario. Ethics approval was not required for this research because it relied exclusively on information that is publicly available. No additional data collection from human participants was required.

Data Analysis

This study used a discourse analysis approach (Potter and Wetherell, 1987) for analyzing the data, which was informed to some extent by conversation analysis (CA) (Sacks, 1992). With its disciplinary home in sociology, CA research has resulted in a large repertoire of empirical work highlighting how interactions—both in text and talk—occur in predictable and orderly ways. Sequences of text and talk have been examined in ways that allow for analysts to consider how the structure of language produces particular social actions. Further, by analyzing how the social actions are produced, suggestions can be made for how to alter language use with the hope of changing practice (i.e., score report use).

As a methodology, CA focuses on micro-analysis of text or talk, attending to how language is sequentially organized. As such, in this study, the tools of CA informed the analysis of score reports. A growing body of research has applied some aspects of CA to the study of text-based data in online contexts (e.g., Lester and Paulus, 2011; Paulus et al., 2016); yet, to date, no research has examined text-based score reports drawing upon this methodological and theoretical lens.

To begin the analysis process, the last author completed an initial line-by-line analysis of the 10 score reports. This initial analysis involved a three-step process. First, this author engaged in repeated readings of the score reports. In discourse analysis work, it is often helpful to begin by familiarizing yourself with the data and to “avoid reading into the data a set of ready-made analytic categories” (Edwards, 1997, p. 89). As such, the initial, repeated readings focused primarily on becoming familiar with the conversational features used within the score reports, as well as their potential functions. Second, this author made analytic memos linked to the score report data that took note of the primary conversation and discourse features employed within the score reports. These memos focused on preparing for a more refined and intensive analysis. Third, the last author developed three broad categories of key conversational features to be considered in greater detail by all three researchers, including: (1) Displays of Information, (2) Knowledge Claims, and (3) Doing Accountability. Within these categories, key conversational features were identified and listed as codes. These codes are presented in Appendix A.

Then, beginning with one student score report, each individual author conducted a line-by-line analysis, applying the coding scheme, while taking note of inconsistencies or challenges in its application. The authors then met to discuss their coded data to ensure consistent interpretation and application of the codes. Then, each author independently coded the remaining data set of nine score reports. After all the score reports were coded, the researchers met again to compare and discuss the coded data, moving to a more fine-grained analysis of key portions of each score report. They took note of differing interpretations, and sought out variability in the data set. Finally, representative extracts were selected and organized around broader patterns/themes, highlighting the implications of language choice. More specifically, in alignment with a discourse analysis approach informed by conversation analysis, explanations/interpretations were formulated around the functions of various discursive features used within the data set. For example, the authors attended to the function of script formulations (i.e., broad, general statements agreed upon by a field) in the score reports, noting how factual knowledge claims made about an individual learner were produced and varied across reports within the testing context. One of the ways by which discourse analysts produce trustworthiness with the reader is through demonstrating arguments by illustrating the steps involved in the analysis of representative excerpts of the data. This often involves a line by line interpretation of the data, showing how participants orient to and work up a given claim. By thoroughly and transparently presenting how each aspect of the findings is supported by excerpts from the larger corpus of

data, the analyst provides space for the reader to evaluate their claims (Potter, 1996). As such, the extracts that were selected are positioned as representative, recognizing that what we share is partial.

Findings

The patterns noted in the present sample of score reports represented micro-discourses embedded within broader discourse related to accountability mandates, such as those associated with NCLB. Throughout the reports ran a thread of communicating to students and families certain requirements of the states and province—to test every child, to report scores, to identify proficiency levels of the student. Claims about a given student largely functioned within a framework of documenting the student's place within established levels of proficiency. This orientation set up the context of language production within the score reports.

Identified Themes and Conversational Features

As noted above, three broad themes were identified within and across the score reports—Displays of Information, Knowledge Claims, and Doing Accountability. Consistent with the perspective of language as action, these themes represented actions generated within the score report texts that reflect the intentions of the educational system within which they were produced. Within each of the broad themes were sub-themes that characterized more fine-grained textual features. In the following paragraphs, the themes and sub-themes are described, presenting examples of representative text drawn directly from the score reports, when applicable.

Displays of Information

In this first theme, language in the score reports functioned through lexical items, typographical features, sequencing, and organization. Together, these elements displayed basic information about a given test. Extract 1 illustrates this well.

Extract 1

Each year, Arkansas students in grades 3–8 take the Arkansas Augmented Benchmark Examination, which assess the Arkansas Curriculum Frameworks and provide national norm-referenced information.

In the above extract, basic information about the Arkansas Augmented Benchmark Examination is presented, with the timeframe (“each year”), who takes the test (“grades 3–8”), and the purpose of the test (“assess...and provide national norm-referenced information”) made explicit. Across the score reports, this type of information was typically placed near the start of the report and functioned to provide specific information about the nature of a given assessment.

Two sub-themes identified were *word choice* and *visual representations*. These sub-themes established the norms of the communication, through tone of the language and emphasis on certain features of a given assessment. The semiotics tradition of communication comes through in this theme. Words in the report function as signs where different subjects in the activity

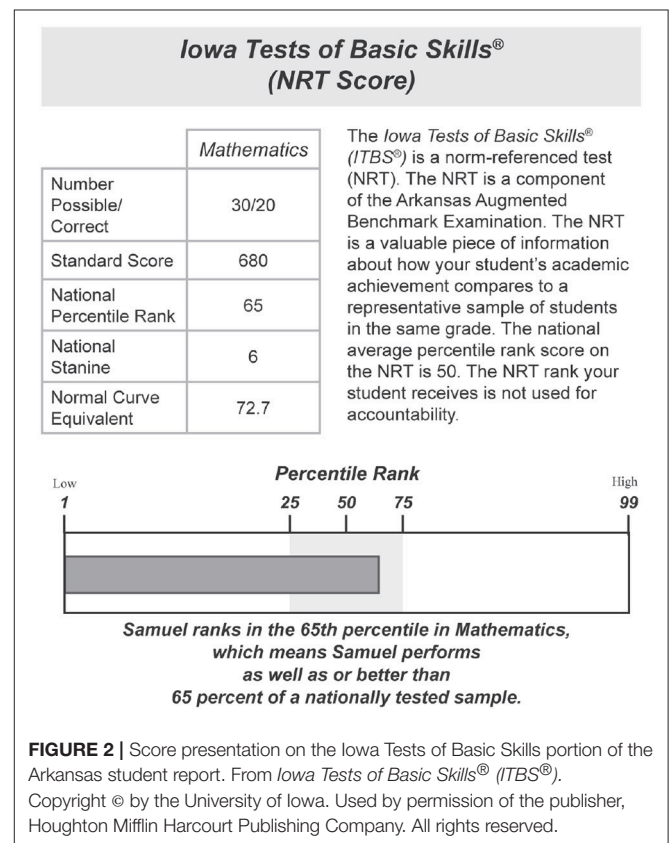
system negotiate the meaning of a student's test performance (i.e., the object). The extent to which these words and other symbolic representations signaling student performance achieve a shared understanding among the subjects dictates their response and how readily they may act toward the overall goal.

Word Choice

In general, the lexical items or word choices created a particular tone in a given report. Chosen words could make the report more personal and conversational or impersonal and formal. For example, use of the child's name throughout a score report conveyed a personal and conversational tone. The individual student was placed as the focus of the scores, whereas repeated uses of “your student” or “your child” functioned to increase the distance between the information within the report and the student's family, reflecting a more impersonal tone.

One of the most notable features around word choice was the use of language regularly used by those working within the testing environment. For example, across the reports various kinds of scores were reported such as scaled scores, stanines, and percentiles. **Figure 2** illustrates this well:

Such jargon (i.e., expert language) is not common language for educational stakeholders such as parents. Further, jargon vs. layperson language often functions to mark language as being designed for only certain audiences or contexts (Housley and



Fitzgerald, 2002). In this case, jargon perhaps marked the score reports as being for an audience familiar with testing language. While definitions for these terms were sometimes provided, placing such testing language throughout the report could serve to reinforce the formality and decrease the accessibility of the report for particular stakeholders.

Proficiency category labels and descriptions represented another example of word choice in action. While these labels were likely determined outside of the report development process, it is within the score report that these labels are communicated to students and families. Several reports featured a label such as *Below Basic* for the category representing the lowest level of performance. Labels associated with higher levels of performance included *Basic*, *Proficient*, and *Advanced*. In the Massachusetts report, however, the lowest proficiency level used the more value-laden label, *Warning/ Failing*. Such language produced a more negative or cautionary tone relative to the more neutral and descriptive label of *Below Basic*. Indeed, the word choices could imply different required actions in response to the report, by varying the perceived sense of urgency related to an individual student's performance.

Likewise, at all levels of performance, proficiency category descriptions associated a given student with a certain academic identity. For example, Extract 2 shows how a student at the highest level of proficiency is positioned as being capable, while a student at the lowest level of proficiency is positioned as inadequate.

Extract 2

Advanced—Students at this level can regularly read above grade-level text and demonstrate the ability to comprehend complex literature and informational passages. Basic—Students at this level are unable to adequately read and comprehend grade appropriate literature and informational passages.

Extract 3 demonstrates how one report emphasized the student's shortcomings, which stands in contrast to Extract 4. In the latter case, language functioned more to describe than to evaluate. Such differences in the words used to describe the student could shape how the student identifies with academics and chooses to proceed based on their test performance.

Extract 3

Below Basic—Students fail to show sufficient mastery of skills in Reading and Writing to attain the Basic level.

Extract 4

Novice—Students state an idea in own writing but provide little or no support and writing contains errors in grammar and mechanics that may affect meaning; have basic knowledge of parts of speech, e.g., verb tense and pronouns.

Visual Representations

Visual representations in the reports were somewhat similar in function to non-verbal elements of oral conversation. In many reports, emphasis of words or phrases within the text were accomplished with the use of visual signaling techniques, such as enlarged font, bolding, and the use of color. The pattern of text that was signaled served to facilitate associations between

the emphasized words or phrases. For example, in **Figure 3**, the Hawaii report created a visual association by emphasizing the student's name and her score relative to the surrounding text. These two pieces of information were given status as the most important content on the page. In contrast, the Iowa report (see **Figure 4**) used large, bold font to indicate the title of the report and student's name, but no strong association is created between the student and any particular score or proficiency categorization.

Related to visual signaling, report content was prioritized with regards to what information was reported, how much space each section of report content was allotted, and the placement of content within the report. Introductory letters at the beginning of some reports implied that establishing a personal connection was important before proceeding to the student's scores. Student scores placed at the top of the page or in another prominent location conveyed or drew the reader's attention to their central importance. Such actions were common across the reports reviewed. Where reports differed was in their treatment of supporting information. Supporting information, such as suggestions for steps to take to help the child succeed, when placed early in the report and in a place toward which the eye easily moves, signaled an importance of interpretation and action on par with identifying the student's score and proficiency level. For example, space was dedicated on the front page of the Ontario report to outline strategies for parents to help support their child's learning based on the test results (**Figure 5**). In contrast, the Arkansas report dedicated a small section to "Additional Resources" on the last page of a four-page report (**Figure 6**). Cases where such information was placed toward the end of the report, in a manner unlikely to attract the eye, or missing from the report altogether, communicated less emphasis on interpretation and action.

Space allotment shaped the tone of the report as well, and, in conjunction with font size in the text, implied a level of importance. For example, a 14-line statement about NCLB policy cast the Missouri report in a bureaucratic light. The report appeared to be as much about fulfilling mandates as it was about identifying the academic progress of the student. A contrasting approach was embodied in the Delaware student report, where the report space was entirely devoted to the student's performance and next steps to be taken (**Figure 7**). An important point about the prioritization of content based on placement and space allotment is that such decisions communicate, either intentionally or not, the relative value of the information from the sender's (i.e., department of education) perspective. Such prioritization could carry a lot of weight, as the sender in this context represents an authoritative body.

Knowledge Claims

The second major theme of language functions centered on conversational features employed when making statements about the student based on their performance on a given test. These statements were worded in such a way that they could be interpreted as factual claims, which, in talk and text, are typically more difficult to challenge. Thus, within the activity system,

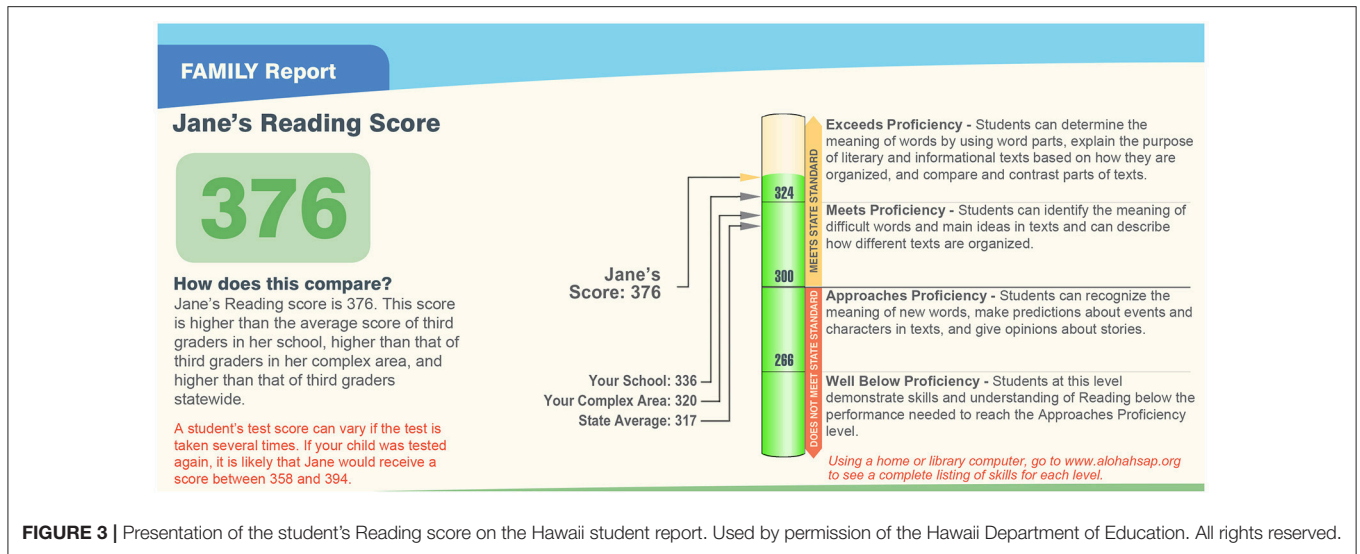


FIGURE 3 | Presentation of the student's Reading score on the Hawaii student report. Used by permission of the Hawaii Department of Education. All rights reserved.

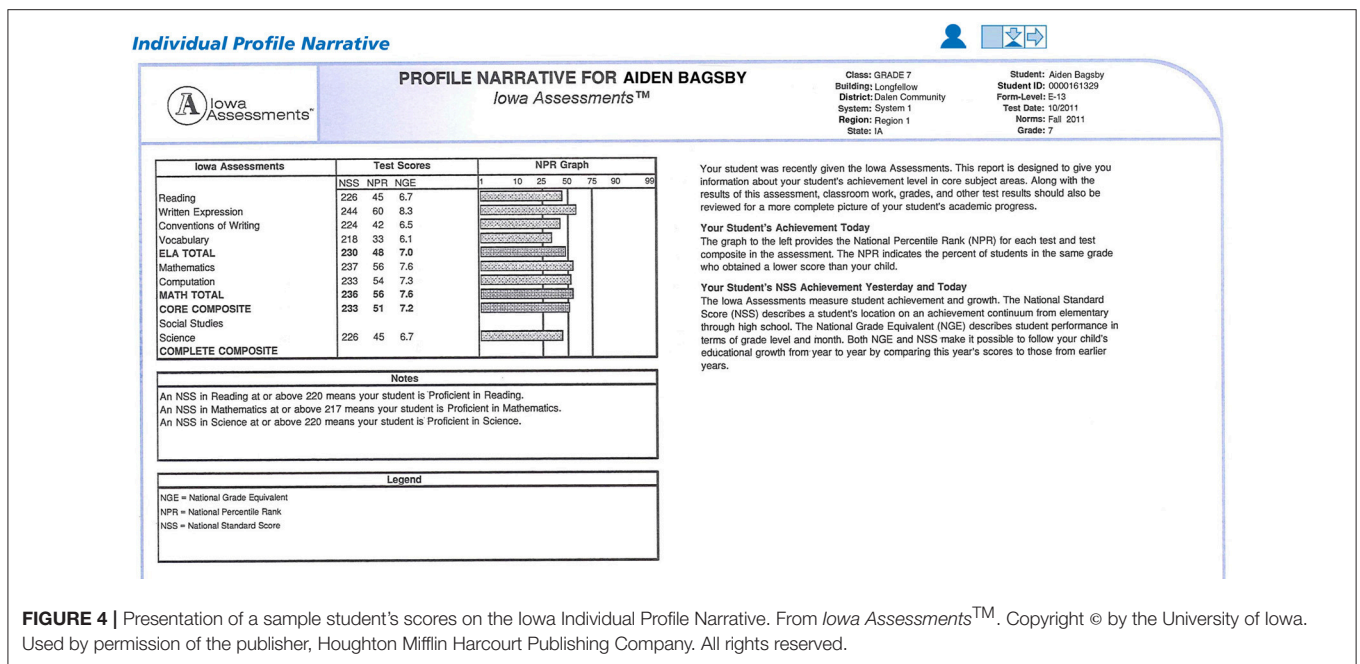


FIGURE 4 | Presentation of a sample student's scores on the Iowa Individual Profile Narrative. From Iowa Assessments™. Copyright © by the University of Iowa. Used by permission of the publisher, Houghton Mifflin Harcourt Publishing Company. All rights reserved.

knowledge claims set up the norms by which the student and their score would be described. They also established educational measurement as the area of expertise that allows for these claims to be made and delimits the strength of the claims. This theme of Knowledge Claims frames the issue at hand in the activity system and how the subjects in the community are to approach this issue. Two sub-themes identified were *script formulations* and *hedging*.

Script Formulations

Script formulations are general statements that often function to present generic knowledge widely accepted as true (Edwards, 1997), and have been discussed extensively in CA literature and other discourse-based studies. Generally, script formulations can

be thought of as broad, general statements agreed upon by a field (e.g., educational measurement and testing), and represent a language feature that could be expected to occur within a score report. Across the score reports, it was noted that script formulations were used when making claims about the student based on the score or proficiency level attained. These statements were phrased generically and could likely be used across students with minimal modifications. Extracts 5 and 6 present examples of script formulations used when reporting percentile scores.

Extract 5

[Student] ranks in the 65th percentile in Mathematics, which means [student] performs as well as or better than 65 percent of a nationally tested sample.



EQAO'S ASSESSMENT OF READING, WRITING AND MATHEMATICS, JUNIOR DIVISION (GRADES 4-6)

Individual Student Report, 2012

EXAMPLE, STUDENT

Ontario education number: **123-456-789**
 School: **Example School (000000)**
 School board: **Example Board (00000)**
 Accommodations: The school requested no accommodations for your child for this assessment.

Your Child's RESULTS

EQAO's junior-division assessment tests the reading, writing and mathematics skills students are expected to have gained by the end of Grade 6. For more information about EQAO assessments and about typical student performance at each level of achievement, see page 2 of this report and "A Parent's Guide to Understanding Your Child's Results," available at www.eqao.com, under "Parent Resources."

	NE 1 Not enough evidence to be assigned a Level 1	Level 1 Much below the provincial standard	Level 2 Approaches the provincial standard	Level 3 Meets the provincial standard	Level 4 Surpasses the provincial standard
Reading: attempted 36 of 36 questions				■	
Writing: attempted 14 of 14 questions				■	
Mathematics: attempted 36 of 36 questions				■	

Each level represents a range of achievement. The position of the ■ shows where, within the range, your child's result is located (from low to high). The shaded line extending from the symbol shows the range of results the student likely would have received if he or she had taken this test or an equivalent test many times.

Your Child's Results on the Primary-Division Assessment (Grade 3, 2009)

Reading Level 3 (meets the provincial standard)

Writing Level 2 (approaches the provincial standard)

Mathematics Level 3 (meets the provincial standard)

Supporting Your Child's LEARNING

Talk to your child's teacher about how these results compare to your child's daily classroom work and assessment information. Discuss how you can work together to support your child's learning.

Here are some suggestions, based on the level your child achieved, that you may find helpful in supporting his or her learning at home.

Reading

- Talk with your child daily about what he or she is reading.
 - Talk about how the people and situations in it are similar to or different from your own experiences.
 - Ask your child for his or her opinion about how situations and problems are resolved in different texts.
- Listen to your child read aloud.
 - Have your child explain how the author uses words and images to make people and events believable and ideas persuasive.

Writing

- Have your child read his or her writing aloud and talk about
 - the words and phrases that are confusing and those that make the meaning clear.
 - how to group ideas into well-developed paragraphs that make the main idea clear.
 - which linking words help the reader follow the sequence of ideas.

Mathematics

- Have your child help you solve everyday problems and talk about how your child can use what he or she is learning in mathematics.
- Play games that require a variety of reasoning skills (e.g., identifying relationships among numbers, grouping and classifying information, estimating, comparing answers to questions).
- Have your child solve a problem in different ways and talk about the advantages and disadvantages of each.
- Talk about the importance of mathematics in everyday activities.



The Education Quality and Accountability Office is an agency of the Government of Ontario.

FIGURE 5 | Strategies to support the child's learning (highlighting added) were presented on the front page of the Ontario Individual Student Report. Used by permission of the Education Quality and Accountability Office. All rights reserved.

Extract 6

The percentile information below compares your child's performance with the scores of students in the same grade across the nation. For example, a student who scores in the 40th percentile performed as well as or better than 40 percent of all students nationally - but not as well as 60 percent of those students.

These statements are to be taken and interpreted as factual, general claims about the student, citing agreed upon definitions and a shared meaning of percentile scores by the educational measurement field.

Script formulations were also used when describing a test's purpose, uses of the results, and descriptions of proficiency categories. These formulations were common

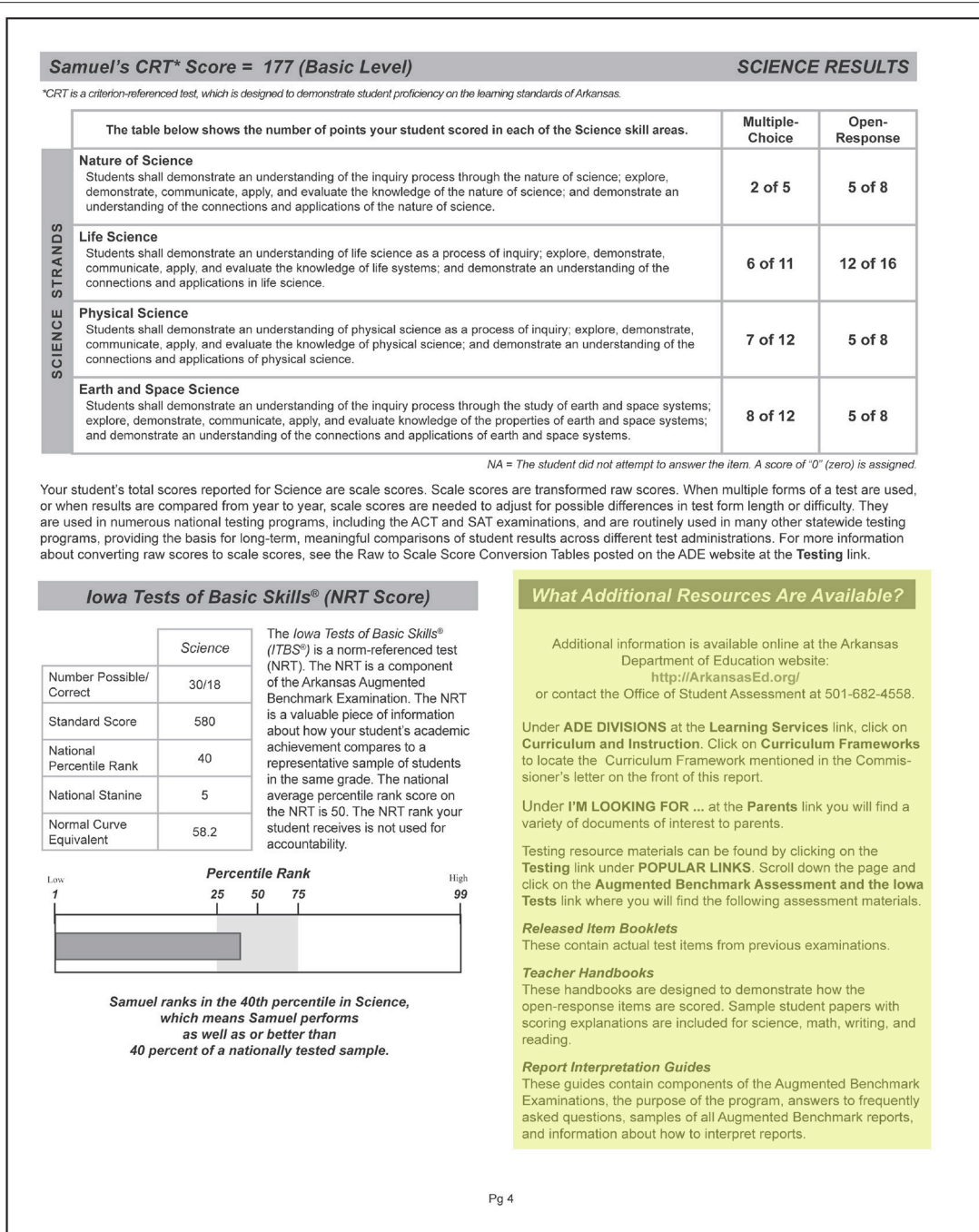


FIGURE 6 | The Arkansas report placed a section on additional resources for the parent (highlighting added) in the bottom corner of the last page. Used by permission of the Arkansas Department of Education. *Iowa Tests of Basic Skills® (ITBS®)*. Copyright © by the University of Iowa. Used by permission of the publisher, Houghton Mifflin Harcourt Publishing Company. All rights reserved.

across score reports, including statements that provided information on the student's level of achievement as measured against set standards, to help inform instruction, or to identify areas in which the student may require additional assistance. Extracts 7 and 8 illustrate such functions well.

Extract 7

This reading test measures your child's performance against the Maryland Reading Content Standards. Maryland's Content Standards describe what all students should know and be able to do.

Susan B. Doe (123456)

Test Date: Spring 2011

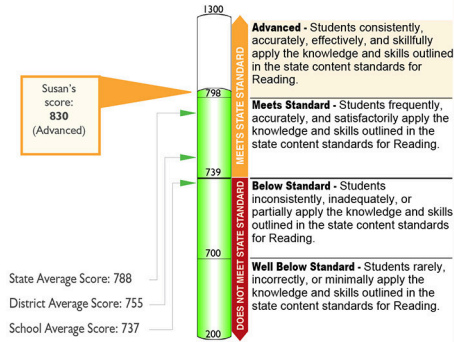
Grade: 5

School: Test School (12345)

District: Test District (12345)

Reading

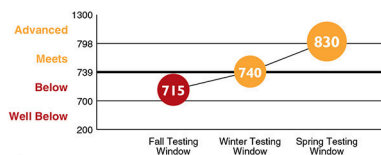
How did Susan do on the Reading Assessment?



Susan's performance on the DCAS Reading Assessment is Advanced, which is estimated to be similar to the Exceeds level on the Delaware Student Testing Program (DSTP).

For more complete Performance Level Descriptors, please visit <http://de.portal.airast.org/educator.html>

How did Susan do on all attempts of the Reading Assessment?



The chart below shows Susan's TerraNova Reading National percentile score compared to other students at this grade.



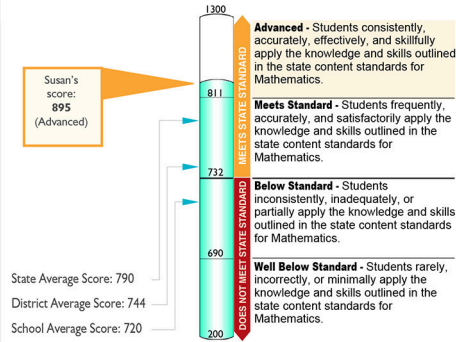
Reading Next Steps

Susan has *met* the state content standards for Reading. However, she may continue to benefit by focusing on general elements of Reading Comprehension and Literary Text. To address Reading Comprehension and Literary Text at home, you could:

- Read a story with your child, and ask her to explain how the point of view of a story impacts the reader.
- Ask your child to explain how the author's decisions (e.g., word choice, content) in an informational text (e.g., newspaper, magazine, product information, biography, editorial) affect the reader.
- Ask your child to explain an author's bias in an informational text by discriminating between facts and opinions.

Mathematics

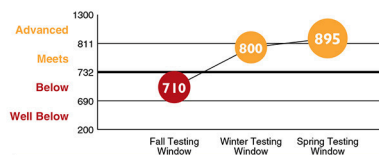
How did Susan do on the Mathematics Assessment?



Susan's performance on the DCAS Mathematics Assessment is Advanced, which is estimated to be similar to the Exceeds level on the Delaware Student Testing Program (DSTP).

For more complete Performance Level Descriptors, please visit <http://de.portal.airast.org/educator.html>

How did Susan do on all attempts of the Mathematics Assessment?



The chart below shows Susan's TerraNova Mathematics National percentile score compared to other students at this grade.



Mathematics Next Steps

Susan has *met* the state content standards for Mathematics. However, she may continue to benefit by focusing on general elements of Mathematics. To address Mathematics concepts at home, you could:

- Have your child calculate the discount for a sale item. Ask her to find the sale price from a percent (e.g., 20% off \$10 = \$2).
- Have her create a table that records the values of a single stock over a week. Ask her to convert this information into a line graph. Ask her to describe the trend or pattern in terms of increases and decreases in value.
- Have her calculate the area of a square, a rectangle, and a triangle. Ask her to explain how these areas are related or different.

For more information about the DCAS (including TerraNova), please visit <http://de.portal.airast.org/>

FIGURE 7 | The space on the Delaware individual student report was dedicated entirely to student performance and next steps to be taken. Used by permission of the Delaware Department of Education. All rights reserved.

Extract 8

Your student's Mathematics Scaled Score is 243 which is at the Nearing Proficiency Level.

Notably, explanations of technical terms were stated in similar ways using script formulations. Extract 9 provides an example from one score report's explanation to assist in visual interpretation of an error bar.

Extract 9

The small gray bar ... shows the range of likely scores your child would receive if he or she took the test multiple times.

Explanations such as this one were common for score report displays of measurement error.

Hedging

In contrast to making factual claims about the student and test, hedging is a conversational feature that mitigates or lessens the impact of a claim, often functioning to distance a speaker/writer from the claim being made and/or to display some form of hesitation or uncertainty (Gabriel and Lester, 2013). Words, such as *but*, *likely*, and *nearly*, and phrases, such as "does not always," work to soften a claim. Across the score reports, hedging words or phrases were commonly used when describing measurement error in practical terms, such as the language presented in Extract 10.

Extract 10

A student's test score can vary if the test is taken several times. If your child was tested again, it is likely that [student] would receive a score between [X] and [Y].

Hedging was also seen in cautionary statements to parents and teachers around using the test as the only indicator of student learning, as seen in Extract 11.

Extract 11

It is important to keep in mind that information on the Student Report is only one source of information about the progress your student is making in school. Grades, classroom work, and other test results should also be reviewed to get a more complete picture of your student's progress.

Con conversationally, the uncertainty introduced by hedging statements related to score precision and use of test results contrasts with the authority of factual claims being made, and could perhaps weaken the claims made throughout the report. Yet, in interactional tasks (both written and spoken) it is common to include such language to soften what could be considered face-threatening (Goffman, 1967). For instance, telling an individual that they performed poorly is a delicate interactional task, as the positive image of their self may be threatened.

Doing Accountability

Another primary theme found in the language of the score reports was communication of roles within the accountability testing context. Within the community of stakeholders engaged in the activity of promoting the student's academic growth, a

chain of authority was produced. Report language established where authority lay and the roles of various entities within the testing context, clearly providing illustration of the lower-right corner of the activity system depicted in **Figure 1**. From the socio-cultural communication frame, language features in the report (re-)establish social orders within the community. Two sub-themes identified were *establishing authority* and *establishing responsibility*.

Establishing Authority

Authority was established in the reports through a variety of language features. Introductory letters often closed with signature blocks that included the chief education officers' middle initials and educational credentials, which positioned the letter's author as possessing the expected membership category or identity for authoritatively writing about the test results. Further, references to NCLB mandates, including appeals to established standards of learning, signaled the report functioned within a high-stakes testing context that required accountability. Extract 12 demonstrates such language.

Extract 12

The Montana Comprehensive Assessment System (MontCAS) was developed in accordance with the following federal laws: Title I of the Elementary and Secondary Education Act (ESEA) of 1994, P. L. 103-382, and the No Child Left Behind Act (NCLB) of 2001.

In the above extract, there is a clear focus on federal legislation, which marks the report as not simply crafted to meet the needs of schools or parents, but also in direct response to legal requirements. Similarly, trademark registrations and test publisher logos, carried a connotation of being associated with larger, authoritative entities. Jargon, such as "augmented benchmark examinations," "curriculum frameworks," and "alignment," also functioned to establish authority, and position the report as being within the confines of a particular group of experts/speakers. In addition to technical terms associated with test scores, as discussed above, jargon could be found in test descriptions, as exemplified in Extract 13.

Extract 13

The North Dakota State Assessment measures essential language arts skills as defined within state standards that are set for the beginning of the current grade level. The assessment is a collection of items carefully selected and reviewed by the assessment publisher, North Dakota teachers, and community members. Test items are weighted and balanced to offer an accurate assessment of what students know and can do. The test offers a meaningful measurement of classroom instruction and student performance. Language Arts scale scores are not used to determine Adequate Yearly Progress for school accountability purposes.

Note how the use of words and phrases such as "essential language arts skills," "items," "weighted and balanced," "Adequate Yearly Progress," and "school accountability" and the appeal to established state standards conveyed the state's test as an authoritative tool within the education system. Across the reports, the names of the assessment programs (e.g., Augmented

Benchmark Examinations) and ubiquitous acronyms (e.g., DCAS, MSA, MontCAS) also projected a sense of authority.

Establishing Responsibility

Language functioned to establish responsibility toward (and away from) three entities—parents, teachers, and the state/province department of education. Introductory letters communicated expected parent responsibilities through statements such as “I invite you to actively participate in [student name]’s education...” and “You should confirm your child’s strengths and needs in these topics by reviewing classroom work, standards-based assessments, and your child’s progress reports during the year.” Section headings, such as “What you can do at home to help your child,” functioned in a similar way. Statements directed toward the parents often worked in conjunction with establishing the responsibilities of teachers.

Across reports, the child’s teacher was identified as a point of contact for interpreting the reports, making sense of the reports in concert with daily classroom work, and working with the parent to support the child’s education. Extract 14 demonstrates one case of establishing such responsibility.

Extract 14

Talk to your child’s teacher about how these results compare to your child’s daily classroom work and assessment information. Discuss how you can work together to support your child’s learning.

Language related to establishing the responsibility of the state/province department of education was complex. The references to NCLB mandates, noted above, conveyed a responsibility on the department to carry out a rigorous testing program to meet federal requirements. Identifying each child’s level of proficiency in the tested subjects was a clear responsibility. There was also language, however, that carried out a careful balance between owning the report and testing program and distancing the department from the claims made about the student. Consider the two excerpts from the introductory letter to the Massachusetts report, presented in Extract 15.

Extract 15

For each test that your child took in spring [year], the report shows your child’s Achievement Level (Advanced, Proficient, Needs Improvement, or Warning/Failing)... If you have questions about your child’s performance, I encourage you to meet with your child’s teacher...

Through the use of the first-person pronoun, *I*, in this and other introductory letters, chief education officers assumed a level of responsibility for the contents reported. One officer even closed her letter with the words, “very truly yours.” She thereby positioned herself as being there to serve the family. On the other hand, key actions such as summarizing the student’s level of academic achievement and placing the student within a proficiency category were cast as actions of not an individual person, but rather of the report or the test. Language such as “this *report* shows” and “this *document* summarizes” [emphases added by the authors] was common and functioned to create distance between an education officer and what was claimed

within the report. In no case did the chief education officer’s letter, for example, include a statement like, “In this report I will summarize your child’s performance...” nor was any such responsibility assigned to the department of education. Pronoun use has been identified as indexing solidarity between speakers/writers and pointing to power differentials between them (Ostermann, 2003). In this analysis, it was noted that first-person pronoun use functioned to mark particular segments of the score report as conveying individual responsibility and power around making particular decisions, with other segments of the score reports deploying third-person pronoun use, which functioned to distance the speaker/writer from the claims being made.

DISCUSSION

The purpose of this study was to examine patterns of language use in score reports within accountability testing in North America. Score reports are an instantiation of language which embodies the socio-cultural context from which they were produced. Activity theory provided the overarching framework for examining score reports as tools (artifacts) within this context, focusing on the social rules, roles, and responsibilities associated with accountability testing. Semiotic and socio-cultural communication perspectives were applied through a discourse analysis methodology (Potter and Wetherell, 1987) informed by conversation analysis (Sacks, 1992). The findings illustrated how micro-conversational features within the score reports are produced within and reflect the broader context (e.g., discourse) of educational testing and accountability. Three themes were identified within the score reports: Displays of Information, Knowledge Claims, and Doing Accountability. These themes represented actions produced within the report and the reproduction of socially constructed expertise. We now proceed to a discussion of each theme, in turn, commenting on the significance of the findings and their implications for score reporting as a mediating artifact within a social system.

The first theme, Displays of Information, concerned the presentation of language used within the reports, including both word choice and visual displays. Among the three themes, this theme has received the most attention in the educational measurement literature. Guidelines such as minimizing the use of jargon and presenting scores in both graphical and narrative forms have been promoted (Goodman and Hambleton, 2004; Zenisky and Hambleton, 2015) in the context of maximizing appropriate interpretations. This orientation to score reporting aligns with an information-processing or cybernetics perspective on communication (Behrens et al., 2013; Gotch and Roduta Roberts, 2018). However, when viewing displays of information through a semiotics communication lens, the choice of words and visual presentation serve to establish tone, and communicate the relative importance of certain types of information. Considering score reports as a relational tool within a social system, this theme suggests that some existing score reporting guidelines could be employed not just to achieve better interpretations, but to facilitate engagement with stakeholders. For example,

minimizing the use of jargon could serve to welcome the audience into the conversation, and the strategic placement of information, such as a greeting addressed to the child in a prominent location on the page, could shift the implied object of central importance from the test score to the child.

The second theme, Knowledge Claims, concerned conversational features used when making statements about a student's performance. As with Displays of Information, the conversational features within this theme—script formulations and hedging—can be viewed in terms of how they facilitate appropriate score interpretations. When reporting a large amount of content to a large stakeholder group, script formulations can represent an efficient and consistent approach to reporting student scores, describing performance, and communicating technical information. Hedging statements, such as descriptions of measurement error, and reminders that the test is but one indicator of a student's learning, can convey foundational understandings of the measurement of latent student attributes. In this way, the statements fulfill mandates specified in professional standards (AERA, APA, and NCME, 2014). From a communication perspective that emphasizes information processes, script formulations and hedging are accurate and unproblematic. However, when viewing script formulations and hedging through a socio-cultural communication frame, the outcome is less clear. As reflections of the cultural norms of the educational community, both conversational features could convey an impersonal tone. At worst, the presence of generic, yet factual-like statements about the child, accompanied by hedging statements that function to soften claims, could confuse, frustrate, and disengage the audience, and create a figurative distance between report users and the test developer or education agency. Conversely, hedging statements that soften claims about the test performance could sustain morale and motivation for students who perform poorly or just miss the cut-off for moving into a higher proficiency level. When viewing score reports as a relational tool, the findings suggest employing communication strategies that balance efficiency and consistency with the conveyance of information about student performance that is enabling and meaningful from the audience's perspective.

The third theme, Doing Accountability, described the implicit communication of roles among those within the accountability testing context. In comparison to the other two themes, this theme has not received much direct attention in the educational measurement literature. Similar to the other themes, the approaches taken in developing the score report, when viewed through a relational lens, bring into focus the social contracts that exist between groups within the educational system. There is an implicit understanding of the roles and responsibilities of students, parents, teachers, and assessment developers. This implicit understanding is communicated within some reporting guidelines. For example, score reports should clearly indicate that the scores are but one indicator of the child's learning and that it should be combined with other indicators of learning within the classroom. While adhering to this guideline, the language within the score report also establishes responsibilities of the parents to follow up with the teacher, and for the teacher

to use the test results to complement their own assessment of the child. Within the score reports reviewed, establishing authority of the testing body was implicitly enacted through the language chosen with the presence of jargon and use of symbols or credentials to signal an authoritative position. In particular, a top-down relationship was established where the test developer, acting on behalf of a federal or provincial governing body, held critical information about the child as well as the privilege to prioritize certain pieces of information, guide the user toward the "right" meaning of the student's test performance, and assign responsibility for acting on the report. Again, while the strategies described thus far in this discussion were employed to facilitate valid interpretations, relationally, they can serve to reinforce a particular social order, specifically one that is assumed rather than negotiated. From this perspective, the implicit communication of responsibility and authority within score reports may have an impact on fulfillment of roles and authentic, sustained engagement with accountability testing by actors within the educational system.

Given the presence of these themes in the reports, it is clear there is a need to consider the development of score reports as well as the noted issues related to their interpretation and use through multiple perspectives. Any actions taken in response to information presented in the score report occurs within the social context of stakeholders interacting with the reports. This social context includes roles and responsibilities and what users define as meaningful in relation to their own needs and expectations. Applying a novel communications frame reveals tensions that may arise in the presence of a disconnect in the meaning different subjects, such as test developers and parents, assign to the test and the student's performance. Such tensions could potentially influence uptake and use of reported information. For example, on the Montana report the text of Extract 12, above, which reflected compliance with federal accountability mandates, is followed with:

MontCAS [Criterion-Referenced Test] scores are intended to be useful indicators of the extent to which students have mastered the materials outlined in the Montana Mathematics, Reading, and Science content standards, benchmarks, and grade-level expectations.

The implication from this passage is clear: the test is intended to measure student mastery against set expectations, and the report will signify how well the student met these expectations. From the perspective of the state department of education and test developer, the primary interpretation of scores should be criterion-referenced, and the test was designed to fulfill a primarily summative function. It has been noted, however, that test score users will often demand diagnostic and instructionally relevant information (Huff and Goodman, 2007; Brennan, 2012). Further, teachers have expressed a need to receive score reports in a timely manner, so they can be responsive in their instructional planning (Trout and Hyde, 2006). Such demands and needs reflect a formative, rather than summative, orientation to the testing occasion. Additionally, Hambleton and Zenisky (2013) stated that examinees are interested in their performance

relative to other test takers, even on criterion-referenced tests. Indeed, further inspection of the Montana report shows these tensions—summative and formative, criterion-referenced and norm-referenced—at play. Even though introductory language clearly establishes a criterion-referenced, summative orientation, additional report content includes the percentages of students across the state scoring within each performance level, and Reading, Math, and Science performances are broken down into strands (e.g., Numbers and Operations, Physical Science).

The source of these tensions can be articulated using the concept of two separate but related activity systems operating within the broader activity system of education. Score reports, as outcomes of one activity system (i.e., test development) reflect the motivations, intentions, and ways of knowing consistent within the educational measurement community. In short, score reports are developed as part of the test development activity system, but are then used as tools by subjects within a different activity system (e.g., classroom learning). In this example, score reports represent a boundary or, more specifically, a *boundary crossing* (Engeström, 2001) between systems. This boundary crossing embodies score reports as common to both activity systems, but tensions may arise regarding their intended use by one system and actual use within another. Thus, boundary crossings between systems provide an opportunity for further study. To-date, investigation of score report use has received little attention relative to stakeholder interpretations. Activity theory can provide a useful framework for addressing this gap in the literature.

Future Directions for Score Reporting Research

Future research in score reporting can begin to further explore boundary crossings between separate but related activity systems within the broader educational system. To extend score reporting research in a socio-culturally oriented direction, future studies could employ ethnographic approaches, studying in-depth the contexts within which score reports are used. The aim of these ethnographic studies would be to develop a thick description of the subjects (Geertz, 1973), their environment, and the interactions between the two, focusing on actions taken in response to the score report. The study of score report use as a socially situated activity may provide an opportunity to document the ways score reports are intentionally integrated by subjects within their own contexts. For example, a study could follow teachers within a school and document their use of score reports and how the reports are (or are not) integrated into teaching practice. This documentation of actual use and practice could then be used to inform a needs assessment that is recommended as a first step in the development of score reports (Zenisky and Hambleton, 2012). In this case, the needs assessment would be informed by a bottom-up approach, grounded in documentation of the users' actual workflow and environment. This approach extends upon proposed methods for conducting an audience analysis which comprises an assessment of audience needs, knowledge, and attitudes (Zapata-Rivera and Katz, 2014). Common among all these methods is the need to

sustain engagement with and understand the target audience. Ethnographic methods are well-suited to provide a richer picture of audience needs, knowledge, and attitudes within their own context, helping to build a literature base that could inform future score report development.

To extend a line of work leveraging an in-depth understanding of user contexts, research and development efforts could consider the potentially divergent values, goals, and assumptions of the testing authority and stakeholders within the reporting context (Behrens et al., 2013). As noted previously, report language, visual presentation, and the implicit prioritization of report content can reflect a bureaucratic orientation. Script formulations and hedging statements, particularly around the communication of measurement error, align with core principles in the measurement field. Both of these conversational characteristics of the reports reflect the intentions and communication goals of the testing authority in the reporting act. What is questionable is whether students and families, the key stakeholder groups, share those same intentions and goals. To the extent that these intentions and goals are not shared, the consequence is potentially failed communication. A need exists for score report research that increases understanding of the goals of stakeholders. Effective score report development would then combine this understanding with a mindfulness of language functions to purposefully select words and displays that negotiate the goals of the testing authority and stakeholders, thus producing mutually beneficial outcomes of the educational system.

Future research studies may also begin to identify the effects of language choices on intended audiences. Report developers should anticipate both cognitive (e.g., readability, understandability) and affective (e.g., motivation, identity formation) factors during the report development process. Lines of research could investigate the dynamics of these factors through experimental manipulations of score report content and layout. In practice, for example, report developers could intentionally attempt to leverage language features that may serve to motivate students whose test performances place them in a “nearing proficiency” category. One form of the report could describe the performance of such students in terms of their deficit in relation to proficiency, while another form could choose language to emphasize the developmental process of learning. Research on formative feedback suggests messages can cultivate positive, learning-oriented goals when they help the learner to see that ability and skills are malleable and influenced primarily by effort, rather than by a fixed aptitude (Hoska, 1993). Well-conducted field testing could assess whether or not such effects on student motivation could be achieved through purposeful language choices in the report. Then the link between student self-beliefs, motivation, and achievement could further extend this line of research.

With regards to roles and responsibilities, findings in the present study point to how teachers may be called upon to partner with parents, and use score reports to improve student performance. Therefore, work that links score reports with expectations and examinations of teacher competency in student

assessment (e.g., Brookhart, 2011; Xu and Brown, 2016) could be useful.

In carrying out work that adopts the proposed approach to score reporting, it is important to remember that the goals and assumptions of the stakeholders will likely shift by testing context (e.g., college entrance, certification). Additionally, different findings may be observed within the accountability context in countries that have experienced a similar push, such as Australia and the United Kingdom. Parallel and complimentary lines of research building on the methodology of the present study will be necessary to build a rich understanding of language functions in score reports.

CONCLUSION

As a prominent source of communication between multiple stakeholders, score reports play an important role within testing systems. We therefore suggest that in addition to issues of interpretability and valid uses in response to score reports, scholars engage in aspects of the communication that are more implicit. Doing so may uncover new approaches for addressing score report use, and push the field to break through complex problems that have persisted throughout the course of score report scholarship. Understanding that language is always *doing*

something, intentionally or not, requires test developers to carefully consider their language choices when communicating with a particular audience. Shifting perspective from the successful transmission of technical information to the role of score reports within a system of activity aimed at improving educational outcomes provides a means for addressing use of the reports and potential social consequences of testing programs. The present study provides a first step to exploring a new dimension within the field of score reporting.

DATA AVAILABILITY

The score report documents and coding notes supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

MR co-conceptualized the research study, participated in data analysis, and led the manuscript writing process. CG co-conceptualized the research study, participated in data analysis, and contributed to the writing process. JL provided methodological support, participated in data analysis, and contributed to the writing process.

REFERENCES

- Alonzo, A. C., de los Santos, E. X., and Kobrin, J. (2014). "Teachers' interpretations of score reports based upon ordered-multiple choice items linked to a learning progression," in *Presented at the Annual Meeting of the American Educational Research Association* (Philadelphia, PA).
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Behrens, J. T., DiCerbo, K., Murphy, D., and Robinson, D. (2013). "Conceptual frameworks for reporting results of assessment activities," in *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (San Francisco, CA).
- Brennan, R. L. (2012). *Utility Indexes for Decisions About Subscores*. CASMA Research Report No. 33. Center for Advanced Studies in Measurement and Assessment, Iowa City, IA.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educ. Meas.* 30, 3–12. doi: 10.1111/j.1745-3992.2010.00195.x
- Craig, R. T. (1999). Communication theory as a field. *Commun. Theory* 9, 119–161. doi: 10.1111/j.1468-2885.1999.tb00355.x
- Edwards, D. (1997). *Discourse and Cognition*. London: Sage.
- Engestrom, Y. (1999). "Activity theory and individual and social transformation," in *Perspectives on Activity Theory*, eds Y. Engestrom, R. Miettinen, and R.-L. Punamaki (Cambridge, UK: Cambridge University Press), 19–38.
- Engestrom, Y. (2006). "Activity theory and expansive design," in *Theories and Practice of Interaction Design*, eds S. Bagnara and G. Crampton-Smith (Hillsdale, NJ: Lawrence Erlbaum), 3–23.
- Engeström, Y. (2001). Expansive learning at work. *J. Educ. Work* 14, 133–156. doi: 10.1080/13639080020028747
- Gabriel, R., and Lester, J. N. (2013). Sentinels guarding the grail: Value-added measurement and the quest for education reform. *Educ. Policy Anal. Arch.* 21, 1–30. doi: 10.14507/epaa.v21n9.2013
- Geertz, C. (1973). *The Interpretation of Cultures: Selected Essays*. New York, NY: Basic Books.
- Goffman, E. (1967). *Interaction Ritual: Essays on Face-to-Face Behavior*. New York, NY: Anchor Books.
- Goodman, D. P., and Hambleton, R. K. (2004). Student test score reports and interpretive guides: review of current practices and suggestions for future research. *Appl. Meas. Educ.* 17, 145–220. doi: 10.1207/s15324818ame1702_3
- Gotch, C. M., and Roduta Roberts, M. (2014, April). "An evaluation of state and provincial student score reports and interpretive guides," in *Paper Presented at the Annual Meeting of the American Educational Research Association* (Philadelphia, PA).
- Gotch, C. M., and Roduta Roberts, M. (2018). *A Review of Recent Empirical Research on Individual-Level Score Reports*. Educational Measurement: Issues and Practice.
- Hambleton, R. K., and Slater, S. (1997). *Are NAEP Executive Summary Reports Understandable to Policy Makers and Educators? (CSE Technical Report 430)*. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Teaching.
- Hambleton, R. K., and Zenisky, A. L. (2013). "Reporting test scores in more meaningful ways: a research-based approach to score report design," in *APA Handbook of Testing and Assessment in Psychology: Vol. 3 Testing and Assessment in School Psychology and Education*, ed K. F. Geisinger, (Washington, DC: American Psychological Association), 479–494.
- Hoska, D. M. (1993). "Motivating learners through CBI feedback: developing a Positive Learner Perspective," in *Interactive Instruction and Feedback*, eds V. Dempsey and G. C. Sales (Englewood Cliffs, NJ: Educational Technology Publications), 105–132.
- Housley, W., and Fitzgerald, R. (2002). The reconsidered model of membership categorisation analysis. *Qual. Res.* 2, 59–83. doi: 10.1177/146879410200200104
- Huff, K., and Goodman, D. P. (2007). "The demand for cognitive diagnostic assessment," in *Cognitive Diagnostic Assessment for Education: Theory and Applications*, eds J. P. Leighton and M. J. Gierl (Cambridge, UK: Cambridge University Press), 19–60.
- Jonassen, D. H., and Rohrer-Murphy, L. (1999). Activity theory as a framework for designing constructivist learning environments. *Educ. Technol. Res. Dev.* 47, 61–79. doi: 10.1007/BF02299477

- Jones, R. C., and Desbiens, N. A. (2009). Residency applicants misinterpret their United States Medical Licensing Exam Scores. *Adv. Health Sci. Educ.* 14, 5–10. doi: 10.1007/s10459-007-9084-0
- Legislative Assembly of Alberta (2017). *Bill 28: School Amendment Act. R.S.A. 2017 (3rd Sess.)*.
- Legislative Assembly of Ontario (1996). *Educational Quality and Accountability Office Act, 1996. S.O. 1996, c. 11*.
- Lester, J. N., and Paulus, T. (2011). Accountability and public displays of knowing in an undergraduate computer-mediated communication context. *Discour. Stud.* 13, 671–686. doi: 10.1177/1461445611421361
- McCall, M. (2016). “Overview, intention, history, and where we are now,” in *The Next Generation of Testing: Common Core Standards, Smarter-Balanced, PARCC, and the Nationwide Testing Movement*, eds H. Jiao and R. W. Lissitz (Charlotte, NC: Information Age Publishing), 19–27.
- National Conference of State Legislatures (2017). *Third-Grade Reading Legislation*. Available online at: <http://www.ncsl.org/research/education/third-grade-reading-legislation.aspx>
- O’Leary, T. M., Hattie, J. A. C., and Griffin, P. (2017). Actual interpretations and use of scores as aspects of validity. *Educ. Meas.* 36, 16–23. doi: 10.1111/emip.12141
- Ostermann, A. C. (2003). Localizing power and solidarity: pronoun alternation at an all-female police station and a feminist crisis intervention center in Brazil. *Lang. Soc.* 32, 251–281. doi: 10.1017/S0047404503323036
- Paulus, T., Warren, A., and Lester, J. N. (2016). Applying conversation analysis methods to online talk: a literature review. *Discourse Context Media* 12, 1–10. doi: 10.1016/j.dcm.2016.04.001
- Potter, J. (1996). *Representing Reality: Discourse, Rhetoric, and Social Construction*. Thousand Oaks, CA: Sage.
- Potter, J., and Wetherell, M. (1987). *Discourse and Social Psychology*. London: Sage.
- Roduta Roberts, M., and Gotch, C. M. (2016, April). “Examining the reliability of rubric scores to assess score report quality,” in *Paper Presented at the Annual Meeting of the National Council on Measurement in Education* (Washington, DC).
- Sacks, H. (1992). *Lectures on Conversation*. Oxford: Blackwell.
- Trout, D. L., and Hyde, E. (2006). “Developing score reports for statewide assessments that are valued and used: feedback from K-12 stakeholders,” in *Presented at the Annual meeting of the American Educational Research Association* (San Francisco, CA).
- United States Congress (2002). *No Child Left Behind Act of 2001*.
- van der Kleij, F. M., and Eggen, T. J. H. M. (2013). Interpretation of the score reports from the Computer Program LOVS by teachers, internal support teachers, and principals. *Stud. Educ. Eval.* 39, 144–152. doi: 10.1016/j.stueduc.2013.04.002
- Washington State Legislature (2017). *High School Graduation Requirements—Assessments Law. ESHB 2224*.
- Whittaker, T. A., Williams, N. J., and Dodd, B. G. (2011). Do examinees understand score reports for alternate methods of scoring computer-based tests? *Educ. Assess.* 16, 69–89. doi: 10.1080/10627197.2011.582442
- Wise, L. L. (2016). “How we got to where we are: evolving policy demands for the next generation assessments,” in *The Next Generation of Testing: Common Core Standards, Smarter-Balanced, PARCC, and the Nationwide Testing Movement*, eds H. Jiao and R. W. Lissitz (Charlotte, NC: Information Age Publishing), 1–17.
- Xu, Y., and Brown, G. T. L. (2016). Teacher assessment literacy in practice: a reconceptualization. *Teach. Teach. Educ.* 58, 149–162. doi: 10.1016/j.tate.2016.05.010
- Zapata-Rivera, D., and Katz, I. R. (2014). Keeping your audience in mind: Applying audience analysis to the design of interactive score reports. *Assess. Educ.* 21, 442–463. doi: 10.1080/0969594X.2014.936357
- Zapata-Rivera, D., and VanWinkle, W. (2010). “Designing and evaluating web-based score reports for teachers in the context of a learning-centered assessment system,” in *Presented at the Annual meeting of the American Educational Research Association* (Denver, CO).
- Zapata-Rivera, D., Zwick, R., and Vezzu, M. (2015). “Exploring the effectiveness of a measurement error tutorial in helping teachers understand score report results,” in *Presented at the Annual meeting of the National Council on Measurement in Education* (Chicago, IL).
- Zenisky, A. L., and Hambleton, R. K. (2012). Developing test score reports that work: the process and best practices for effective communication. *Educ. Meas.* 31, 21–26. doi: 10.1111/j.1745-3992.2012.00231.x
- Zenisky, A. L., and Hambleton, R. K. (2015). “A model and good practices for score reporting,” in *Handbook of Test Development, 2nd Edn*, eds S. Lane, M. R. Raymond, and T. M. Haladyna (New York, NY: Routledge), 585–602.
- Zwick, R., Zapata-Rivera, D., and Hegarty, M. (2014). Comparing graphical and verbal representations of measurement error in test score reports. *Educ. Assess.* 19, 116–138. doi: 10.1080/10627197.2014.903653

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Roduta Roberts, Gotch and Lester. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX A

CODING STRUCTURE USED TO IDENTIFY CONVERSATIONAL FEATURES WITHIN SCORE REPORTS

Displays of Information

Structure and use of lexical items, typographical features, sequencing, and organization to display basic information within a report

Sub-themes:

Word Choice: Lexical items or word choices that created a particular tone in a given report

Visual Representation: Visual signaling techniques and placement of report content to convey relative emphasis—similar to non-verbal elements of oral conversation.

Knowledge Claims

Conversational features employed when making statements

about the student based on their performance on a given test

Sub-themes:

Script Formulation: General statements that often function to present generic knowledge widely accepted as true

Hedging: A conversational feature that mitigates or lessens the impact of a claim, often functioning to distance a speaker/writer from the claim being made and/or to display some form of hesitation or uncertainty.

Doing Accountability

Communication of roles within the accountability testing context

Sub-themes:

Establishing Authority: Production of a chain of authority that positions groups into various roles, and signifies exclusivity of certain groups.

Establishing Responsibility: Statements that established responsibility toward (and away from) three entities—parents, teachers, and the state/province department of education.