Elementary Students Justifying Explanations and Reasoning with Evidence in Academic Tasks and Non-academic Routines

Alison L. Bailey (UCLA), Sandy Chang (CRESST/UCLA), Anne Blackstock-Bernstein, Eve Ryan, & Despina Pitsoulakis (UCLA)

Copyright © 2015 UC Regents of the University of California

The views expressed in this report are not those of U.S Department of Education, the Wisconsin Department of Public Instruction, WIDA or UCLA/GSE&IS/CRESST. All errors remain our own.
Objectives

In this paper we focus on elementary students’ use of evidence to support their own thinking. Previous studies of evidence in persuasive explanations have typically focused on older students, (e.g. middle and high school), examined students’ performance with academic tasks only (e.g., science or mathematics), and/or required students to choose which story characters provide the best justifications and explain the reasoning behind their choices (e.g., Sandoval & Cam, 2010).

In contrast, this study examined young students’ justifications for their own actions in explanations of both academic tasks (mathematics) and personal routines (teeth cleaning).

Specifically, we investigated (1) whether explanations of everyday routines and academic procedures differ in their requirements for evidence, (2) how justifying explanations and reasoning with evidence may differ over time, and (3) whether justifying explanations and reasoning with evidence may be related to English language arts and mathematics test performances.

Perspectives

Explanation is a form of school-relevant language critical for academic success (Nippold & Scott, 2010; Christie, 2012). Students’ abilities to explain are important to understand not least because they are fundamental to child development; early explanatory talk has been linked to cognitive, literacy, and later, discourse development (e.g., Snow, 1991). In developmental studies, explanation is found to be important for revealing “underlying casual relations and properties” to young children (Legare, 2012, p.183).

In academic contexts, much research has been directed at the role of explanations in children’s acquisition of new knowledge. In the area of English language arts, Goldman and Wiley (2011) used discourse analysis to understand 6th grade students’ understanding of texts.
and found that “causal self-explanation during reading was related to better reasoning scores” (p. 127). In the field of mathematics education, explanations have been defined as “giving mathematical meaning to ideas, procedures, steps, or solution methods” (Hill, Charalambous & Kraft, 2012, p. 63). In terms of children’s use of evidence in their understanding of the domain of science, Sandoval and Cam (2010) found that 3rd and 4th grade students preferred definitive data as a source of evidence as opposed to plausible causal mechanisms or appeals to authority, although when ambiguous data was presented, students preferred plausible causal mechanisms as a source of evidence. Children’s reasons for preferring a given justification were related to its perceived credibility, which children often associated with the first-hand experience (related by others, in this case by fictional characters) involved in the data collection process.

Explanation is a productive language practice found to be common across academic content area discourse, texts, and tests (Bailey, Butler, Stevens, & Lord, 2007). We also focus on explanation because of its importance as a key language function cutting across the academic domains of the college and career ready standards (CCRS; English Language Proficiency Development Framework, Council of Chief State School Officers [CCSSO], 2012). Use of explanations to demonstrate understanding in the classroom has also been shown to be an effective metacognitive strategy that can promote student learning and problem-solving (e.g., Aleven & Koedinger, 2002). In the era of new standards, even young students must be able to explain their understanding, and teachers are urged to employ student explanations to gauge new knowledge (e.g., National Governors Association Center for Best Practices & CCSSO, 2010).

In terms of arguing from evidence, abilities to provide reasoned judgements accompanied by evidence may be capabilities that are honed in educational settings and have been found to be acquired by 9th grade. Kuhn (1991) for example, suggested that children’s ability to construct
arguments may be related to their participation in formal schooling, specifically the "academic’ discourse mode may encompass the attitude that assertions must be justified and that alternatives should be considered” (p. 290). Osborne and Patterson (2011), however, make the critical distinction between explanation and argument as discursive practices—“arguments are essential to the process of justifying the validity of any explanation as there are often multiple explanations for any given phenomenon” (p. 629). Consequently, whether elementary students can combine these discursive practices to aptly display their reasoning with evidence, still needs to be determined. Moreover, following Kuhn, on the one hand, we might expect to find that providing an argument to justify one’s explanation of a chosen mathematical procedure may be more accessible to children because of the support of argument construction in a formal schooling context. On the other hand, children may find that providing an argument for a familiar personal routine (e.g., teeth cleaning) is more accessible to them because they are well acquainted with the content and are expert at their own personal routines.

Method

Participants:

We elicited a series of oral explanations from 40 students during their 5th grade school year. The sample (see Table 1) was drawn from a larger project (n=324) studying K-6 grade students’ oral and written language development in the context of academic and non-academic tasks (Bailey & Heritage, 2014). The main aim of the larger study was to establish a trajectory or progression of the features of oral and written explanations from a linguistically diverse sample of students in order to inform teaching. The 5th grade participants in the current study were in four of the five participating schools. Informed consent was obtained from parents of all participating students. Demographic data were requested from the school/district and from
students’ teachers in order to obtain students’ gender, birth date, English language learner (EL) status, and academic performance data; either the California Standards Test (CST-English Language Arts and Mathematics) or the Stanford Achievement Test 10 (Total Reading and Total Mathematics).

Table 1
Student demographic and academic performance information (N=40)

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>20 (50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>20 (50)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EL status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English learner</td>
<td>11 (27)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English only/proficient</td>
<td>29 (73)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age in Years,Months</strong></td>
<td>40 (100)</td>
<td>10;8</td>
<td>0;5</td>
<td>10;1-11;5</td>
</tr>
<tr>
<td><strong>CST Mathematics scale score</strong></td>
<td>20 (50)</td>
<td>383</td>
<td>89</td>
<td>241-511</td>
</tr>
<tr>
<td><strong>CST ELA scale score</strong></td>
<td>20 (50)</td>
<td>354</td>
<td>39</td>
<td>266-418</td>
</tr>
<tr>
<td><strong>SAT10 Total Math scale score</strong></td>
<td>19 (47)</td>
<td>687</td>
<td>50</td>
<td>617-812</td>
</tr>
<tr>
<td><strong>SAT10 Total Reading scale score</strong></td>
<td>20 (50)</td>
<td>679</td>
<td>33</td>
<td>622-732</td>
</tr>
</tbody>
</table>

*One student had a missing SAT 10 Total Math scale score.

**Procedure:**

The oral explanation data used in the current study were generated using two oral elicitation tasks at two time points approximately 4-6 months apart (henceforth T1 and T2). The language and cognitive demands associated with the tasks are deliberately designed to be decontextualized, requiring students to explain their chosen processes to a hypothetical student who is not present. First, the student had to make this realization, take account of the limited point of view of the hypothetical student, and reflect this understanding linguistically in their attempts to be fully explanatory. Two sessions were conducted one-on-one with a researcher in a quiet room or hallway on school grounds during the school day. They were audio recorded and
transcribed verbatim. At the first session, the researcher began with a warm-up activity, in which the student was presented with an illustration of a child doing homework, and asked a series of questions about how and why the student does his/her homework. Following the warm up, the researcher presented one of the two tasks. In the case of the teeth cleaning task, the student was shown an illustration of a gender neutral child cleaning his/her teeth and told, “This is a picture of a girl/boy cleaning her/his teeth. Now, I’m going to ask you a few questions. Please give your best explanation for each one.” Students then responded orally to a series of prompts that culminated in the elicitation of explanations analyzed for this study, specifically “Pretend you’re talking to a friend who doesn’t know how to clean his/her teeth. When you’re ready, tell him/her how to do it and why he/she should do it.” ¹

A second task given to the students one-to-two weeks later was designed to elicit language for mathematical understanding as it is reflected in the Common Core State Standards for Mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010).² Specifically, the students were expected to interact with mathematical concepts of counting and cardinality and express understanding of mathematical procedures through oral explanations (Bailey, 2013). Students were presented with a quantity of colored cubes (plastic interlocking blocks) and asked to find the total number of cubes (see Figure 1; 100 cubes for 5th grade). Students were told to find the total using whatever method they wished, and these strategies were characterized (Bailey, Blackstock-Bernstein, & Heritage,

---

¹ Pronouns referring to the classmate were matched to participant’s gender.
² The Counting and Cardinality Kindergarten CCSS Mathematics Standards 4 & 5 which read in part “Understand the relationship between numbers and quantities; connect counting to cardinality…” and “Count to answer ‘how many?’ questions about as many as 20 things arranged in a line, a rectangular array, or a circle, or as many as 10 things in a scattered configuration; given a number from 1–20, count out that many objects” (National Governors Association Center for Best Practices, CCSSO, 2010).
In this subsample of the larger study, half of the students used addition (50%) as their strategy during the mathematics task, followed by counting (30%) and multiplication (20%). After providing an answer, students were asked to respond orally to the prompt “Pretend you are talking to a classmate who has never done this activity. When you're ready, tell him/her how to use the cubes to find out how many there are and why using the cubes this way helps him/her.”

Figure 1. Participant chooses to array cubes, implements multiplication strategy and is prompted for explanation of procedures (photograph used with permission)

Coding and analysis

Analysis of the 153 explanations elicited during the two tasks at the two time points was conducted using a coding scheme based on prior work by Fitzgerald and Baird (2011) for characterizing the nature of the students’ reasoning and in a part on Sandoval and Çam (2011) to describe the nature of evidence the students may have included. The coding decisions were made
by consensus process for 14 students’ transcripts for both math and social routines at T1 and T2. This procedure led to the refinement of the scheme before the authors each coded six additional students (independent reliability on 24 (16%) additional transcripts is being determined; challenging coding decisions encountered by coders continued to be resolved by consensus of the research team). The final coding scheme captures whether students provided justifications, if they needed to be prompted for justification of their chosen procedures during the course of giving the explanations, and if the justification was comprehensible (logical) to the coder, in addition to characterizing the reasoning or argument and the nature of the evidence (see Appendix).

**Results**

Students almost always provided justification for their explanations although approximately a third required additional prompting for justification of their explained personal routines at both time points and for math explanations at T1 only (see Table 2). Most students’ justifications were logical (e.g., cause and stated effects made sense) given their explanations of their teeth cleaning routine and their chosen math strategy (i.e., their procedures for counting, addition or multiplication).

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>Personal Routine Explanations T1 n (%)</th>
<th>Personal Routine Explanations T2 n (%)</th>
<th>Math Explanations T1 n (%)</th>
<th>Math Explanations T2 n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanation includes justification</td>
<td>40 (100)</td>
<td>39 (100) *</td>
<td>40 (100)</td>
<td>34 (100) †</td>
</tr>
<tr>
<td>Prompted for justification</td>
<td>11 (28)</td>
<td>12 (31)</td>
<td>12 (30)</td>
<td>5 (15)</td>
</tr>
</tbody>
</table>
Logical justification | 39 (98) | 36 (92) | 35 (88) | 30 (88)  
Reasoning types:‡ | | | | |
Stated facts | 12 (30) | 7 (18) | 8 (20) | 7 (21)  
Causal function | 8 (20) | 7 (18) | 6 (15) | 3 (9)  
Causal action | 9 (22.5) | 8 (20.5) | 17 (42.5) | 15 (44)  
Normative analogy | 0 (0) | 2 (5) | 0 (0) | 0 (0)  
Normative authority | 1 (2.5) | 4 (10) | 1 (2.5) | 0 (0)  
Normative “first principles” | 31 (78) | 24 (62) | 30 (75) | 23 (68)  
Other ψ | 2 (5) | 3 (8) | 2 (5) | 0 (0)  
Evidentiary types: ‡ | | | | |
No evidence | 36 (90) | 37 (95) | 27 (68) | 24 (71)  
Perception | 2 (5) | 1 (3) | 2 (5) | 2 (6)  
Likelihood | 1 (2.5) | 0 (0) | 2 (5) | 1 (3)  
Hearsay | 0 (0) | 0 (0) | 0 (0) | 0 (0)  
Comparative | 0 (0) | 1 (3) | 7 (17.5) | 8 (23.5)  
First-hand experience | 1 (2.5) | 1 (3) | 3 (7.5) | 2 (6)  

* One student did not give a justification for the explanation; valid percent based on n=39.  
† One student did not give a justification for the explanation and five students did not have a T2 math explanation; valid percent based on n=34  
‡ Does not sum to total number of students because more than one type can be used per student explanation  
Ψ Predominantly also normative in nature but not captured by other normative categories

First, we considered whether explanations of everyday personal routines and academic procedures such as a math problem-solving strategy differed in terms of their reasoning and their requirements for evidence at the two time points.

**Personal routine explanations**

Most students gave just one distinct form of reasoning to justify their personal routines (21 and 24 students, respectively at T1 and T2), 14 students at both time points gave two different types of reasoning and five students at T1 and just one at T2 had three different types. These justifications were characterized as predominantly *normative* reasoning, appealing to “first
principles” (Fitzgerald & Baird, 2011) for why students would need to clean their teeth well (e.g., hygiene, cleanliness). The next most favored reasoning was causal action, which focused on needing to clean teeth (in the manner explained by the students) in order to keep them white, stop cavities, etc. (see Table 3 for examples). Very little evidence was provided with students’ justifications for their personal routines. Indeed 36 students at T1 and 37 students at T2 provided no evidence for their justification for cleaning their teeth, and four provided just one type of evidence at T1, and at T2 two students provided one type of evidence, and just one student provided two types of evidence.

Math explanations

The number of different argument types provided for math explanations at T1 was divided between students providing just one (n=19) or providing two (n=18). Just three students provided three different types of reasoning. At T2, fewer students gave multiple types of arguments in justifying their math task procedure, with most providing just one argument type (n=21) and fewer giving two (n=12) and three (n=1) types of arguments, While these justifications were characterized as predominantly normative reasoning, as with personal routines, appealing to “first principles” (e.g., efficiency, parsimony), almost half the sample provided causal reasoning (e.g., gave cause and effect actions) for their chosen math problem-solving procedures (see Table 2).

The proportion of students who omitted evidence was also lower with the math explanations than with the personal routines at both time points (see Table 2). In contrast with personal routine explanations, just 27 students at T1 and 24 students at T2 provided no evidence for their justification for their chosen math procedure. Twelve provided at least one type of evidence at T1, and at T2 seven students provided one type of evidence. Just one provided two
types of evidence at T1 and three provided two types at T2. The favored type of evidence to emerge in the math explanations at both time points was *comparative* (see Table 3 for examples).

**Table 3**
Example excerpts of different types of reasoning and evidence used to justify explanations

<table>
<thead>
<tr>
<th>Justification by reasoning type</th>
<th>Examples (with evidentiary link, if included)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement of fact</td>
<td>“And people recognize colors more than random little cubes all sorted out in different ways.” (T2) [Nascent evidentiary link (i.e., still no specific observation or experimentation cited for the claim): people’s <em>visual perception</em>]</td>
</tr>
<tr>
<td>Causal claims</td>
<td>“And then if you clean your teeth, it will make your teeth brighter and more healthier because when your teeth are like yellow, they’ll be like more weak.” (T2) [No evidence, such as “My dad had a yellow tooth that went bad”]</td>
</tr>
<tr>
<td></td>
<td>“Because it'll make it a lot easier instead of counting every single cube altogether. So it's better if you make it with, with the same color of piles to make it easier if because if you put all the colors together, then if you just keep on counting one by one, whichever order, you might count count the same one twice, and you might get the wrong answers. It's better because piles make it more organized by the same colors or numbers. And it'll make it a lot more organized so that's why.” [Evidentiary link: <em>comparison</em> between two math problem-solving strategies; <em>visual perception</em> (largely nascent - no specific experimentation cited for the claim)]</td>
</tr>
<tr>
<td>Normative (by subtype):</td>
<td></td>
</tr>
<tr>
<td>Analogy</td>
<td>“Because it's the same as the, as like cleaning your body except it's in your body.” (T2) [No evidence, such as “I got sick when I didn’t keep my body clean”]</td>
</tr>
<tr>
<td>Authority</td>
<td>“...and your dentist won't be mad at you....” [No specific observation cited for the claim]</td>
</tr>
<tr>
<td>“First principles” (e.g.,</td>
<td>“...it’s good for you to brush your teeth so you don’t get judged by bad breath or like not so good teeth.” (T1) [No evidence, such as people’s visual and olfactory perceptions and dislikes based on these]</td>
</tr>
<tr>
<td>efficiency/parsimony for math; oral hygiene or aesthetics/appearance for teeth cleaning)</td>
<td></td>
</tr>
</tbody>
</table>
“Because using the cubes this way is the best way because you can easily use your multiplication skills to find it instead of having to add them all together, which would probably take a long time.” (T2) [Nascent evidentiary link (still no specific observation or experimentation cited for the claim): prior first-hand experience with two different math problem-solving strategies]

Even though we see patterns of difference between the two explanation tasks (less so between time points) with such a small sample and indeed so few students providing evidence, we cannot meaningfully conduct statistical tests of significance to determine whether personal routines and mathematics explanations differ in terms of type of reasoning given in their justifications and their evidence usage and type.

We also considered whether justifying explanations and reasoning with evidence may be related to English language arts and mathematics test performances, as well as given the linguistically diverse sample, English language learner status. We converted students’ ELA and mathematics scale scores to z-scores in order to create single scales for ELA and mathematics of the different academic performance tests available for the students. Students’ abilities to give a logical justification in math explanations at T1 was significantly related to their performance on their ELA examination, $F(1, 38) = 4.96, p = .03$. No additional relationships between the measures of academic performance scores and EL status and providing justification and evidence were statistically significant.

**Discussion and Conclusions**

Overwhelmingly, students relied on normative arguments to justify both personal routine and math explanations at both time points—appealing to “first principles” (Fitzgerald & Baird, 2011)
such as imperatives for healthy teeth and dental aesthetics, or imperatives to optimize efficiency and strive for accuracy in math problem-solving, often when they also provided causal arguments in their explanations.

Rarely did students use evidence; surprisingly even explanations of the familiar personal routine tended not to cite evidence. As we consider Kuhn’s (1991) suggestion that the formal educational situation may pull for argumentation in students, we see tentative evidence that it was within the math explanations that we saw students making more cause and effect based arguments and that these explanations were the main site of what little evidentiary bases we found in the explanations overall.

Specifically, in this sample of 5th grade students, the beginnings of evidentiary links occurred most often in math explanations where many children went beyond the normative assertions dominating the personal routine explanations to also provide causal reasoning for their actions and to also invoke evidence from comparisons between the math strategies they knew. However, these emergent uses of an evidentiary basis for arguments were largely still absent any overt references to specific observations that would have relied on students’ visual perception or other forms of data (i.e., overtly stated prior first-hand experiences such as experimentation with different math strategies) that would perhaps constitute more adequate (convincing) evidence.

Conceivably, empirical evidence was not readily invoked in either elicitation context. The math task specifically may have generated more justification of the student’s chosen problem-solving strategies had we reversed the order of the task (i.e., told the students there were 100 cubes in the pile and asked them to prove it). This reversal may provide more context and scope for overt mathematical reasoning. Alternatively, students may think justifications based on theories about causal mechanisms or social norms/assertions are sufficient for a naïve listener.
The current context may have been too simplistic and students may have made assumptions about the knowledge state of the fictitious friend/student they were make-believing was the recipient of their explanations or made assumptions about the knowledge state of the actual recipient, the adult researcher, or both.

The current sample is small and may explain why only one significant finding for academic performance (ELA) was linked to student ability to give a logical justification in an oral explanation task. No additional relationships with either ELA or mathematics performance are surprising but scale scores on standardized academic assessments could be too distal a measure of the kinds of language and mathematics abilities being drawn upon for the explanation tasks. In next steps, we will increase sample size for the analyses (conducting power analyses to determine this), include teacher ratings of the mathematical understanding evident in the explanations as well as measures of the quality of the language used in the explanations themselves. Nippold (2009) in a study of 7-15 year olds hypothesized that greater domain/content knowledge may be associated with better language abilities in students’ explanations. This hypothesis did not hold and Nippold speculated that this was due to an exclusive focus on the measurement of the syntactic quality of the explanations, whereas students can still form syntactically complex sentences for content they know little about or have an imperfect understanding of. Therefore our future work will include measures of language that go beyond both distal ELA assessment and the syntactic structure of explanations to include measures of discourse quality such as the coherence (organization across sentences) and cohesion (connections between linguistic elements within and between sentences) of student oral explanations that may prove sensitive to (and thus can fall victim to) student’s content knowledge.
While the preliminary results can only be suggestive, they caution that not all explanations may entail an articulation of evidence. Rather teachers need to be aware that even as students justify their explanations, the companion discourse practice—argument—may need additional instructional support if students are to include cogent reasoning, and along with this, an articulation of pertinent evidence (Sandoval & Millwood, 2005).
Acknowledgements

We would like to thank the students and teachers who participated in this research and the school administrators and staff who worked with us. We acknowledge the many contributions of our colleagues, research assistants and programmers, but especially wish to thank Margaret Heritage of WestEd and Barbara Jones of CRESST. Alison Bailey acknowledges funding from the WIDA Consortium at the Wisconsin Center for Educational Research and the ASSETS Enhanced Assessment Grant from the U.S. Department of Education. However, the contents do not necessarily represent the policy of the U.S. Department of Education and you should not assume endorsement by the federal government. Alison Bailey acknowledges serving as a consultant and advisory board member for WIDA projects.

References


Appendix

Coding Scheme

Step 1: Does the child provide a justification? Yes/No

Step 2: Did they need to be prompted/reminded? Yes/No

Step 3: Is the justification comprehensible (makes sense/logical)? Yes/No

Step 4: What kind of reasoning or argument is in the justification? (rate all that apply)
   a. Stated facts (but nothing linking these facts together; just co-vary; correlation)
   b. Causal: chained order of facts, actions, organized causal claims – mechanisms by
      which the world works (one action entails another; need a temporal order)
      i. Causal function - how objects in the tasks (explained procedure) work
         (i.e., how teeth cleaning equipment functions to arrive at clean teeth; how
         cubes function to arrive at answer).
      ii. Causal action - actions in the explained procedure (i.e., by brushing
         teeth well; by counting by 2s, 5s)
   c. Normative: own values - examined critically by working out foundational/first
      principles that must be “defensible and logically connected to “the normative
      contention” (F&B, 2011, p.623).
      i. Uses analogy (i.e., cleaning teeth is important like washing hair everyday)
      ii. Appeals to an authority (i.e., teacher told students to count this way =
          strategy)
      iii. Appeals to “first principles” for the specific domain (e.g., Math task:
          Need to optimize efficiency & strive for accuracy is important; Oral
          hygiene: Need to stay healthy & aesthetics/appearance is important)
   d. Others?

Step 5: What kind of evidence is in the justification? (rate all that apply)
   i. no evidence
   ii. Visual/olfactory perception (i.e., I saw it worked..., my teeth then smelt)
   iii. Likelihood (i.e., you are very sure to lose your teeth if you don’t clean
       them)
   iv. Hearsay - (i.e., My dad already had one [cavity], and he told me how bad
       it feels: pretty bad. So I brush my teeth)
   v. Comparative outcomes (i.e., doing it this way is faster than counting by
      1s)
   vi. First-hand experience (i.e., it helped me this way so it will help you;
       ....because it lets me keep my teeth healthy for eating)