A formative assessment of students' algebraic variable misconceptions

Joan Lucariello a, b, Michele T. Tine c, b, *, Colleen M. Ganley d, b

a The City University of New York, United States
b Boston College, United States
c Dartmouth College, United States
d Florida State University, United States

ABSTRACT

Gaining an accurate understanding of variables is one challenge many students face when learning algebra. Prior research has shown that a significant number of students hold misconceptions about variables and that misconceptions impede learning. Yet, teachers do not have access to diagnostic tools that can help them determine the misconceptions about variables that their students harbor. Therefore, a formative assessment for variable misconceptions was created and administered to 437 middle- and high-school students. Analyses from the test scores were found to exhibit strong reliability, predictive validity, and construct validity in addition to important developmental trends. Both teachers and researchers can use the test to identify students who hold misconceptions about variables.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

Learning algebra is a "gatekeeper" to students' future educational and career success (Adelman, 2006; RAND Mathematics Study Panel, 2003; Silver, 1997; U.S. Department of Education, 1999). An increasing number of school districts have responded recently by adding algebra to their high school graduation requirements (Achieve, 2007). Given its importance, it is disquieting that learning algebra proves so challenging. Data from the National Assessment of Educational Progress (NAEP) show that algebra achievement of U.S. students is poor, with only 6.9% of 17-year-olds scoring at or above a proficient level (National Center for Educational Statistics, 2005).

One significant problem is that many students experience difficulty mastering foundational algebraic concepts, one of which is an understanding of variables (Knuth, Alibali, McNeil, Weinberg, & Stephens, 2005; Kuchemann, 1978; Philipp, 1992). Moreover, misconceptions (alternative conceptions) about variables are common among students (e.g., Kieran, 1992; Kuchemann, 1978; Rosnick, 1981; Stacey & Macgregor, 1997). Yet, diagnostic assessments about variable misconceptions are not available to teachers. Therefore, the primary goal of the current study is to develop an assessment with reliable and valid items that can specifically diagnose if students harbor misconceptions about variables. The secondary goals are to (1) determine how common misconceptions about variables are among middle- and high-school students and (2) explore developmental trends in the formation of misconceptions about variables.

To understand some of the typical misconceptions that students hold about variables, it is best to begin by charting correct understanding of variables. Proper understanding of symbols as variables includes a few key components. First, the variable

* Corresponding author at: Department of Education, Dartmouth College, 6103 Raven House, Hanover, NH 03755, United States. Tel.: +1 603 646 9043.
E-mail address: michele.tine@dartmouth.edu (M.T. Tine).

0732-3123/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmathb.2013.09.001
must be interpreted as representing an unknown quantity. That is, a student must realize that a symbol represents a unit that does not have an ascertained value. Second, a student must interpret the symbol as representing a varying quantity (Philipp, 1992) or range of unspecified values (Kieran, 1992). This is known as the “multiple values” interpretation of literal symbols (Knuth et al., 2005). These first two proper interpretations have been studied by presenting seventh and eighth graders who have been exposed to curriculum about variables with the problem: “The following question is about the expression ‘2n + 3.’ What does the symbol (n) stand for?” (Knuth et al., 2005). Correct responses expressed the idea that the literal symbol (1) represents an unknown value (e.g., “the symbol is a variable, it can stand for anything”) and (2) could represent more than one value (e.g., “it could be 7, 59, or even 363.0285”). However, approximately 39% of seventh graders and almost 25% of eight graders gave incorrect responses (e.g., “I don’t know” or “nets” or “5”). These data provide clear evidence that a sizable group of students do not correctly interpret variables.

A third component of understanding variables entails awareness that some kind of relationship exists between symbols as their value changes in a systematic manner (e.g., as $h$ increases, $r$ decreases) (Kuchemann, 1978). Said differently, a correct interpretation of variables entails knowing that related numbers that change together are “variables.” (Philipp, 1992). The “which is larger” problem has been relied on to assess this understanding. For example, Kuchemann (1978) presented 3000 high-school students who had been taught about variables with the following problem: “Which is the larger, $2n$ or $n + 2$? Explain.” Only 6% of students were correct and seemingly aware of a “second order relationship,” that the relation between $2n$ and $n + 2$ is actually changing with $n$. Indeed, the difference between $2n$ and $n + 2$ increases as $n$ increases. When $n = 2$, the two expressions are equal: when $n = 3$, $2n > n + 2$. Knuth et al. (2005) also explored this understanding, using the “which is larger” problem with middle-school students. Only about 18% of sixth graders, just over 50% of seventh graders, and 60% of eighth graders evidenced the understanding that a relationship exists between symbols because their value systematically changes.

However, the issue is not as simple as students lacking correct knowledge about variables. Of additional concern is that many students actually hold erroneous concepts about variables. Often students come to school with knowledge of concepts in the curriculum. If this knowledge is inconsistent with the concepts being taught, the knowledge is termed an alternative conceptions or misconceptions (Lucariello, 2009). Considerable research has documented that many misconceptions in mathematics and science are quite common. The current study focuses on three of the common, major misconceptions about variables that students experience (as described in the literature) and develops an instrument that detects these three misconceptions.

The first of these misconceptions was initially documented by Kuchemann (1978) during an exploration of students’ interpretations of variables. Specifically, he found some students consistently ignored variables. For example, in the problem “Add 4 onto $n + 5$,” 68% of students answered correctly ($n + 9$), while 20% of students gave the incorrect answer $9$, suggesting they simply ignored the variable $n$ altogether.

A second type of misconception is seen when students treat variables as a label for an object (McNeil et al., 2010). This was shown by Stacey and Macgregor (1997) when they presented more than 2000 middle school students the following problem: “David is 10 cm taller than Con. Con is $h$ cm tall. What can your write for David’s height?” The correct answer is $10 + h$, wherein 10 is added to the number or quantity denoted by $h$. Yet many students treated the variable as a label associated with the name of an object (e.g., $C + 10 = D$). Based on other research findings, interviews with individual students, and coding of students’ informal or written explanations, Stacey and Macgregor (1997) interpreted this answer to reflect ‘$C$ as meaning ‘Con’s height’ and $D$ as meaning ‘David’s height’). Another similar erroneous concept is seen when students interpret the variable as an abbreviated word (e.g., response of $D h$ where the abbreviation stands for the words David’s height).

This misconception of construing a variable as a label for an object is reflected also in the classic error to the “Students and Professors” problem, which reads as follows: “Write an equation, using the variables $S$ and $P$ to represent the following statement. ‘At this university there are six times as many students as professors.’ Use $S$ for the number of students and $P$ for the number of professors.” An erroneous understanding that $S$ is a label for an object (students), as opposed to a variable (number of students), led 37% of a sample of students entering college to incorrectly answer the question as $6S = P$ (Rosnick, 1981). When asked to explain this answer, students stated that they believed the answer was $6S = P$ because $S$ was a label for students. (The correct answer is $S = 6P$ where $S$ stands for number of students.) This misconception reasoning on this “student and professor” problem was prevalent also among students already in college (Clement, Lochhead, & Monk, 1981). Another example of the misconception of a variable as a label for an object/entity is seen when students, who are given the question “In the expression $t + 4$, what does represent?,” answer with “time” instead of “any number”.

Finally, a third type of misconception is when students believe a variable is a specific unknown (Kuchemann, 1978; Stacey & Macgregor, 1997). In this case, students do not fully understand that a variable can represent multiple values, rather they believe it can only represent one fixed value. For example, when asked how many values $p$ represents, students assume $p$ can only hold one value, as opposed to many values. This contradicts the correct understanding of a variable previously discussed.

Misconceptions are particularly important for teachers to know about, as misconceptions can impede learning. The process of student learning varies contingent on whether students’ preinstructional knowledge of a given concept(s) accords (or not) with correct curricular concepts (concepts in the domain). When student preinstructional knowledge is correct and consistent with correct curricular/domain knowledge, student knowledge is conceived of as “anchoring conceptions.” When preinstructional knowledge is incorrect and hence runs contrary to what is being taught, student knowledge is
described as “misconceptions” (or “alternative conceptions”). In cases where students have misconceptions, learning is more challenging (see Gelman & Lucariello, 2002 for review; Hartnett & Gelman, 1998). If a student holds a misconception, the misconception interferes with or distorts the assimilation of other inputs, such as the correct concepts presented in a given curriculum. In these cases, learning is a matter of conceptual change or accommodation, wherein current knowledge must undergo substantial reorganization or replacement (Carey, 1985, 1986; Posner, Strike, Hewson, & Gertzog, 1982; Strike & Posner, 1985, 1992). Learning as conceptual change is more difficult than learning as conceptual growth. Conceptual growth occurs when one’s anchoring conceptions (previously held knowledge that is consistent with new input) provide a base for assimilating curricular inputs, leading to the establishment of new concepts or enrichments of current ones. One example of conceptual growth would be when preinstructional knowledge about counting principles serves as a base for learning about addition (Hartnett & Gelman, 1998). There is nothing about verbal or nonverbal counting that is inconsistent with the idea that addition involves putting together two sets. However, as noted, sometimes preinstructional knowledge is inconsistent with new concepts and conceptual change is needed. The counting principles are inconsistent with the mathematical principles underlying the numberhood of fractions (Hartnett & Gelman, 1998). One cannot count things to generate a fraction and one cannot use counting based algorithms for ordering fractions. For instance, \( \frac{1}{4} \) is not more than \( \frac{1}{2} \). In order to learn about fractions, conceptual change is needed as a student would need to reorganize or replace their previous theory about counting.

As a result of conceptual change being difficult to accomplish, researchers have suggested specific instructional strategies to help teachers bring it out in their students. While traditional methods of instruction, such as lectures, labs, discovery learning, and reading text, can be effective at achieving conceptual growth, they are generally ineffective at bringing about conceptual change (Chinn & Brewer, 1993; Kikas, 1998; Lee, Eichinger, Anderson, Berkheimer, & Blakeslee, 1993; Smith, Maclin, Grosslight, & Davis, 1997). Instead, conceptual change requires particular instructional strategies, such as raising student metacognition and creating cognitive conflict by creating experiences during which students consider their erroneous knowledge side by side with or for the same time as the correct concept or theory (see Lucariello, 2009; Mayer, 2008). Because overcoming misconceptions requires conceptual change and conceptual change is not easily achieved, misconceptions tend to be entrenched in student reasoning (Brewer & Chinn, 1991; McNeil & Alibali, 2005). The change-resistance account argues that children’s misconceived base knowledge can easily become reinforced through uncorrected erroneous practice and as a result their ability to learn more complex tasks within a given domain suffers increasingly over time (McNeil & Alibali, 2005).

For all of these reasons, it is critical for teachers to know about misconceptions their students might have about variables. Yet, resources and tools to assist them in uncovering such misconceptions are not readily available for teachers in the United States. Currently, tests that are freely available for teachers to use are typically of little diagnostic value. For example, required state and/or national standardized tests may assess variables, but are summative in nature and their purpose does not include informing individual instruction (Educational Testing Service, 2012). There are a handful of general diagnostic mathematics tests available for teachers to purchase (e.g., The Stanford Diagnostic Mathematics Tests, Group Mathematics Assessment and Diagnostic Evaluations, KeyMath Diagnostic Assessments, The Diagnostic Test of High School Math, and The Diagnostic Test of Pre-Algebra Math). While these some of these tests include a few items about variables, they do not specifically test for, diagnose, or report on the nuanced and distinct types of variable misconceptions. The algebra version of the Chelsea Diagnostic Mathematics Tests (see Hart, Brown, Kerslake, Küchemann & Ruddock, 1985) appears to be the only preexisting test that includes at least one item for each of the common misconceptions discussed in the current study. However, even in this case, the scoring and reporting is not done in a way that helps teachers identify which item(s) map to which variable misconception.

The goal of the current study is to develop a test with items that are reliable and valid that can be used by teachers to diagnose the misconceptions about variables that students may have. The information on student thinking gained from such a test could be used to guide appropriate instruction and thereby facilitate student learning as conceptual change.

2. Method

2.1. Participants

This study included two pools of participants totaling to 483 students. Participants were omitted from the analyses if they did not complete all nine items on the diagnostic test \( n = 24 \), if they were in Pool 1 and did not complete all of the problems in the isomorph set \( n = 20 \), or if they listed their Grade as 13 \( n = 2 \). This left 437 total participants from the following grades: 6th \( n = 4 \), 7th \( n = 70 \), and 8th \( n = 145 \), 9th \( n = 108 \), 10th \( n = 64 \), 11th \( n = 36 \) and 12th \( n = 10 \). Pool 1 included 217 participants and Pool 2 included 220 participants. Also see Table 1.

To describe students with respect to their algebra course experience, four levels of algebra experience were defined. Students with No Formal Experience had never taken a formal Algebra course; Students with Minimal Experience were currently taking an Algebra 1 course; Students with Basic Experience had completed an Algebra I course only; Students with Moderate Experience had completed an Algebra 1 course and were either taking or had already completed an Algebra 2 course. Please note that it is likely that all students, even those in the No Formal Experience group had informal exposure to algebraic concepts during elementary school.

The gender distribution across Pools 1 and 2 was 57% boys and 43% girls and 60% boys and 40% girls, respectively. Of the 205 Ss and 203 Ss from Pools 1 and 2, respectively, that chose to identify their race, the majority were White (64%; 53%),
with another 11% and 13% African American or Black, 10% and 11% Hispanic or Latino, 5% and 6% Multiracial, 2% and 6% Asian or Pacific Islander, 1% and .5% Native American, and 6% and 10% Other.

Chi Square analyses showed that subject pools did not differ from each other in the distribution of participants across grade ($\chi^2 (6, n = 437) = 9.328, p = 0.16$), algebra course experience ($\chi^2(3, n = 437) = 1.759, p = 0.62$), gender ($\chi^2(1, n = 437) = 0.495, p = 0.48$), or race ($\chi^2 (6, n = 437) = 9.328, p = 0.21$).

Students came from classrooms for which the teacher had enrolled the class to participate in this online study of student algebraic reasoning. Teachers were recruited nationally from advertisements with various teacher-membership associations. The recruitment advertisements were identical for teachers of all grade levels and the response rate for participation was similar across grade levels. In exchange for participation, teachers received information on their students' performance on the algebra ability test that was administered.

### 2.2. Procedure

#### 2.2.1. Data collection procedures

Data were collected online during intervals of the school day designated by the respective teachers. Measures included a misconceptions diagnostic test, cognitive (near and far) transfer problems, a student background questionnaire, and an algebra ability test and were administered online in the order listed. Each is described in detail in the measures section below. Once students completed any one measure, they no longer had access to that measure.

#### 2.2.2. Reliability and validity analysis procedures

The reliability of the misconceptions diagnostic test scores was tested in two ways. First, Cronbach's alphas were calculated for students' correct and misconception responses on the diagnostic test. Cronbach's alpha is a coefficient of internal consistency and is commonly used as an estimate of the reliability of a psychometric test for a sample items (Allen and Yen, 2002). Second, the correlation between the responses on the diagnostic test items and the near transfer test items was calculated. This was done in order to try to understand the extent to which students generalize their algebraic knowledge about the original test items (in terms of correct, misconception, and incorrect responding) to items identical in structure, but distinct in surface features (see subsection on near transfer problems below for the specific ways the items differed from those on the diagnostic test).

The validity of the misconceptions diagnostic test scores was also tested in two ways. First, to examine predictive validity, a hierarchical multiple regression analysis was run which determined the extent to which misconception responding can predict an external criterion (specifically, Algebra Ability Test performance). Second, the construct validity of the diagnostic test items was determined by examining students' transfer of their correct, incorrect, and misconception responses on the diagnostic test to the near and far transfer problems. Specifically, it was expected that transfer levels would be above chance for correct and misconception responding, reflecting that correct and misconception understandings are entrenched forms of thinking. Meanwhile, it was expected that the transfer levels for incorrect responding would be at or below chance, as incorrect responding should not necessarily reflect a systematic or conceptually based form of thought.
2.3. Measures

2.3.1. Misconception diagnostic test

In order to create test items for the diagnostic test, a number of items were developed and pilot tested online with a sample of 313 students. These items consisted of released multiple-choice items from international, national, and state level mathematics assessments, such as TIMSS and NAEP. Next, diagnostic test items were written and adapted from these piloted items in order to specifically assess variable misconceptions. Therefore, each item was designed so that there were four answer options in which one indicated a misconception, one was the correct choice, and two were incorrect choices (foils). In contrast to the misconception answer choice, the incorrect answer choices were simply wrong answers, and did not represent any specific misunderstanding about variables. The test was made up of a total of nine multiple-choice items (see Appendix). Three items (1–3) were designed with a misconception response that reflected the erroneous understanding that a variable can be ignored. Three items (4–6) were designed with a misconception response that reflected an erroneous understanding that a variable is a label for an object. Three items (7–9) were designed with a misconception response that reflected the erroneous interpretation that a variable was a specific unknown, as opposed to one that can hold varying values. Responses within each group of items were significantly correlated with one another, with $p < 0.05$ in all cases. The nine items were presented to each participant in a random order.

2.3.2. Cognitive transfer problems

Each pool of participants completed a set of cognitive transfer items to help assess the likelihood that students generalize their knowledge (in terms of correct, misconception, and incorrect responding). Pool 1 participants completed 9 isomorph (near transfer) items and Pool 2 participants completed 3 non-isomorph (far transfer) problems.

Near transfer problems. The nine near transfer isomorph problems were constructed to each match a corresponding item in the diagnostic problem set. Each isomorph problem was identical to its match problem in the diagnostic set in problem structure, vocabulary, and order of answer options. It varied from its match by a change only in the actual letters and numbers used in the problem (see sample in Fig. 1). Thus the same algebraic understanding would be required to correctly answer each item as for the corresponding diagnostic test item. Similarly, the misconception answer choice on the original diagnostic test and the isomorph test indicate the same misunderstanding about variables.

Far transfer problems. Three non-isomorph problems were constructed, one each to correspond with each of the three ways of misinterpreting a variable. Each problem was constructed as a variant of one of the 3 diagnostic test problems about each of the three misunderstandings. These problems differed from their diagnostic set mate on three features (more features than the isomorph problems differed from their respective diagnostic set mate). Each varied from its match by changes in the numbers used, sign of the numbers (positive/negative), and arithmetic operation (see sample in Fig. 1). The far transfer problems were identical to their match problems in the diagnostic set in problem structure, vocabulary, and order of answer options.

![Sample diagnostic test item and corresponding transfer items.](image-url)
Table 2
Hierarchical regression analysis of variables predicting the proportion of correct responses on the algebra ability test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B (SE)</th>
<th>β</th>
<th>R²</th>
<th>ΔR²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−0.02 (0.02)</td>
<td>−0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algebra Course Experience D1</td>
<td>0.07 (0.02)</td>
<td>0.17***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algebra Course Experience D2</td>
<td>0.06 (0.03)</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algebra Course Experience D3</td>
<td>0.00 (0.03)</td>
<td>0.00</td>
<td>0.04***</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number incorrect responses</td>
<td>−0.04 (0.005)</td>
<td>−0.35***</td>
<td>0.22***</td>
<td>0.18</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number misconception responses</td>
<td>−0.05 (0.006)</td>
<td>−0.35***</td>
<td>0.33***</td>
<td>0.11***</td>
</tr>
</tbody>
</table>

Note: Betas are for the final step in the model.
* p < 0.05.
** p < 0.01.
*** p < 0.001.

2.3.3. Student background questionnaire
Demographic and background information (e.g., experience with algebra courses) of students was collected through this questionnaire.

2.3.4. Algebra ability test
This test consisted of 20 multiple-choice questions that measure algebra knowledge culled from different administration dates of Grade 10 Mathematics Massachusetts Comprehensive Assessment System (MCAS) exam. The 20 items were specifically selected because they did not include incorrect option choices that reflected common mathematical misconceptions (i.e., the misconceptions focused on in this study as well as other misconceptions not associated with variables, like those associated with equality). One test item was omitted due to administration problems, therefore students’ proportion of correct responses on the Algebra Ability Test was calculated as the total correct out of 19 items ($α = 0.72$).

3. Results

3.1. Reliability of the diagnostic test

To assess the internal consistency of the diagnostic test items, Cronbach’s alphas were calculated based on both students’ correct and misconception responses on the diagnostic test. Ideally Cronbach’s alpha should be above 0.7, although Cronbach’s alphas of 0.6 or higher are often acceptable during test construction and especially on tests that do not have a large number of items (Cortina, 1993; DeVellis, 2003). For correct responding, the Cronbach’s alpha was 0.77, indicating that this set of items measured a single unidimensional latent construct, algebraic understanding, to a high degree. For misconception responding on the diagnostic test, the Cronbach’s alpha was lower, 0.46, which makes sense when considering the test was specifically designed to test distinct misconception types.

As mentioned above, students in Pool 1 ($n = 217$) also completed an alternate form of the diagnostic test, the isomorph problem set. As another way of assessing the reliability of the scores on the diagnostic test, the correlation between the responses on the diagnostic test and the isomorph test was ascertained. The correlations between the responses on these two tests were 0.85 for proportion of correct responses, 0.71 for proportion of incorrect responses, and 0.49 for misconceptions responses. It is important to note that the correlations for the proportion of incorrect responses may be higher than the proportion of misconception responses because the proportion of incorrect responses measures the relationship between choosing either of the two incorrect options on the diagnostic tests and either of the two incorrect options on the isomorph tests, while the proportion of misconception responses measures the relationship between choosing the one misconception option on the diagnostic test and the one misconception option on the isomorph test. The mean incorrect answer rate on the diagnostic test was 0.18 ($SD = 0.18$) and 0.19 ($SD = 0.17$) on the isomorph test. The mean misconception answer rate on the diagnostic test was 0.22 ($SD = 0.13$) and 0.20 ($SD = 0.14$) on the isomorph test.

Because the correlations between the diagnostic and isomorph tests were quite high, Cronbach’s alphas were calculated for correct and misconception responding for the diagnostic test items and the isomorph test items together – thus there was a total of 18 items in the scale. The reliability for correct responding increased to 0.88 and for misconception responding, the reliability increased to 0.66. The results of these analyses suggest that the diagnostic test items are reliable.

3.2. Validity of diagnostic test

3.2.1. Predictive validity
The extent to which misconception responding predicts an external criterion (Algebra Ability Test performance) was examined. To do this, a hierarchical multiple regression analysis was run (Table 2). The proportion of correct responses on
the Algebra Ability Test was the dependent variable. In Block 1 gender and 3 dummy variables for algebra course experience were entered as covariates. In Block 2 the number of incorrect responses on the diagnostic test was entered. In Block 3 the number of misconception responses on the diagnostic test was entered. The results of this analysis showed that the number of misconceptions students had explained an additional 11% of the variance in Algebra ability test scores over and above gender, algebra course experience, and incorrect responding ($F(1, 357) = 60.081, p < 0.001$).

### 3.2.2. Construct validity

Construct validity of the diagnostic test items was determined by examining students’ transfer of their correct, incorrect, and misconception responses on the diagnostic test to near and far transfer problems. Specifically, it was expected that transfer levels would be above chance for correct and misconception responding, reflecting that correct and misconception understandings are entrenched forms of thinking. Meanwhile, it was expected that the transfer levels for incorrect responding would be at or below chance, as incorrect responding should not necessarily reflect a systematic or conceptually based form of thought.

To assess transfer of correct responses, a student was assigned a 1 if they transferred a correct response on the diagnostic set problem to the corresponding near transfer problem. The student received a 0 on transfer if a correct response on the diagnostic test did not transfer to the near transfer problem, but instead a misconception or incorrect response was made on the transfer problem. Transfer of incorrect and misconception responses were calculated in the same way. These transfer levels were then compared to chance level. This was done because students were twice as likely to answer with an incorrect response since there were two incorrect answer options compared to only one misconception and one correct answer option, as explained in the reliability analysis section above. In other words, due to the number of answer choices that represented a correct response (1/4), a misconception response (1/4) and an incorrect response (2/4), chance level responding for misconception and correct responding (25%) was different from chance for incorrect responding (50%).

On near transfer problems, students transferred their correct knowledge at a much higher rate than chance (58% above chance), their misconceptions at a rate a bit higher than chance (20% above chance), and their incorrect knowledge at approximately chance level (2% below chance). On far transfer problems, students transferred their correct responses at a rate of 55% above chance, their misconceptions at 12% above chance, and their incorrect responses at 3% below chance. These data, which showed transfer levels above chance for correct and misconception responses, both presumed to be conceptually-based, supports the construct validity of the diagnostic test items.

Another way to examine construct validity is to examine how the Pool 2 participants performed on the problems that were most different from the diagnostic problems. Specifically, Pool 2 participants completed three far transfer items, which differed from the diagnostic test items more than the isomorph items. The correlations between the scores on the diagnostic test items and scores on the far transfer items were 0.73 for proportion of correct responses, 0.56 for proportion of incorrect responses, and 0.42 for misconceptions, thereby providing further evidence of construct validity.

### 3.3. Overall student performance on diagnostic test

#### 3.3.1. Descriptive

On average students got 5.59 items correct (62%) and answered 1.73 items with a misconception response (19%) and 1.68 with an incorrect response (19%) (Note: chance level responding for incorrect responding was 50% for incorrect responding and 25% for correct and misconception responding). Twelve percent of students (59 students) received a perfect score on the test. Tables 3 and 4 show descriptive statistics and frequencies, respectively, for the different response types on the diagnostic test.

#### 3.3.2. Item-level analyses

Table 5 shows the percentage of students giving different response types on each of the nine test items. As can be seen, the nine diagnostic test items differed drastically in the proportion of students who chose the misconception response. Between 7% and 44% of students selected the misconception response on a particular item.

---

1 Incorrect responses were chosen to be entered in Block 2 as opposed to correct responses because correct and misconception responses are highly correlated ($r = -0.67$) compared to the correlation for incorrect and misconception responses ($r = 0.17$). Therefore, including both correct and misconception responses into a regression analysis together presents multicollinearity issues. When both entered, tolerances were 0.45 and 0.46 for correct and misconception responses, respectively, and VIFs were 2.24 and 2.19 for correct and misconception responding, respectively.
Table 4
Frequencies of different response types on the diagnostic test.

<table>
<thead>
<tr>
<th>Number of Responses</th>
<th>Correct</th>
<th>Misconception</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>109</td>
<td>135</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>108</td>
<td>111</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>98</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>65</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>32</td>
<td>38</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>53</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Analyses were run to determine if high-misconceiving students were more likely than other students to choose misconception responses broadly, that is, regardless of an item’s tendency to elicit (or not) misconception responses. If high-misconceiving students choose misconception responses even on “easy” problems, or those that are less likely to elicit misconception responses from the sample in general, it would underscore how entrenched misconception reasoning tends to be. Items were grouped into two categories – those with a high percentage of misconception responding (3 items ranging from 20% to 44% misconception responding) and those with a low level of misconception responding (6 items ranging from 7% to 18% misconception responding). Then students were grouped based on their misconception responding on the 9-item test. Students who had above chance misconception responding (3 or more misconception responses; \(N = 122\), 28% of students) were categorized as Misconceivers and all other students were classified as not being misconceivers.

A repeated measures ANOVA with Misconception Amount (High/Low) as a within-subject variable and Misconceiver Group (Misconceiver/Not a Misconceiver) as the between-subject variable was run. The dependent measure was the number of misconceptions. Results showed a main effect of Misconceiver Group, \(F(1, 435) = 634.10, p < 0.001\) but no interaction between Misconceiver Group and Misconception Amount, \(F(1, 435) = 3.32, p = 0.07\). Thus, on both items that were more likely to induce misconception responding and on those that were less likely to induce misconception responding, Misconceivers had much higher misconception responding than other students. This suggests that regardless of problem difficulty, Misceivers were more likely to select misconception responses.

3.3.3. Types of misconceptions
To assess whether certain types of misconception responses were more common and whether Misceivers and other students differed in their most common misconception types, a repeated measures ANOVA with Misconception Type (variable as a label, variable as specific unknown, variable ignored) as a within-subject variable and Misconceiver Group (Misceivers/Not a Misceivers) as the between-subject variable was run. The dependent measure was mean number of misconceptions. There was a main effect of Misconception Type, \(F(2, 435) = 35.66, p < 0.001\), such that the most common type of erroneous understanding was that the variable is a label for an object (25%), followed by the erroneous interpretation that a variable is a specific unknown (19%), as opposed to one that can hold varying values, and the least common was to ignore the variable (13%). There was no interaction \(F(2, 435) = 0.538, p = 0.58\), suggesting that this pattern of performance held for both Misconceivers and other students.

3.3.4. Developmental trends in variable misconceptions
To understand whether there is a developmental change in students’ misconception responding, an analysis of variance was conducted with grade level (middle school or high school) as the independent variable and number of misconceptions on the diagnostic test as the dependent variable. Middle school participants included the 219 participants from 6th (\(n = 4\)), 7th (\(n = 70\)), and 8th (\(n = 145\)) grades. High school participants included the 218 participants from 9th (\(n = 108\)), 10th (\(n = 64\)), 11th (\(n = 36\)) and 12th (\(n = 10\)) grade. Results show that high school students (\(M = 0.24, SD = 0.17\)) had more misconceptions than middle school students (\(M = 0.16, SD = 0.10\)).

Table 5
Percentage of students giving correct, misconception, and incorrect responses on the nine diagnostic test items.

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Correct</th>
<th>Misconception</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>56%</td>
<td>17%</td>
<td>27%</td>
</tr>
<tr>
<td>2</td>
<td>70%</td>
<td>14%</td>
<td>16%</td>
</tr>
<tr>
<td>3</td>
<td>72%</td>
<td>9%</td>
<td>18%</td>
</tr>
<tr>
<td>4</td>
<td>76%</td>
<td>14%</td>
<td>9%</td>
</tr>
<tr>
<td>5</td>
<td>39%</td>
<td>44%</td>
<td>17%</td>
</tr>
<tr>
<td>6</td>
<td>45%</td>
<td>18%</td>
<td>37%</td>
</tr>
<tr>
<td>7</td>
<td>79%</td>
<td>7%</td>
<td>14%</td>
</tr>
<tr>
<td>8</td>
<td>64%</td>
<td>20%</td>
<td>16%</td>
</tr>
<tr>
<td>9</td>
<td>57%</td>
<td>30%</td>
<td>13%</td>
</tr>
</tbody>
</table>

Note: There were two incorrect answer choices and only 1 correct and 1 misconception answer choice.
4. Discussion

The primary aim of the present research was to develop a formative assessment with scores that are reliable and valid and thus able to provide teachers with insight into misconceptions their students hold about variables. These aims were achieved. The instrument was successfully created. Its reliability was proven in two ways. First, the scores demonstrate strong internal consistency. In addition, the correlation of responses across the diagnostic test and an alternate form of the diagnostic test were high for correct and incorrect responses and moderately high for misconception responses.

The scores on the diagnostic test were also shown to be valid. Predictive validity was shown through a hierarchical multiple regression analysis. The number of misconception responses on the diagnostic test significantly predicted student performance on an algebra ability test comprised of items from the Massachusetts Comprehensive Assessment System tests. Specifically, the number of misconception responses a student had explained an additional 11% of the variance in ability test scores above and beyond gender, algebra course experience, and the number of incorrect responses. These results show that after taking into account variables known to be related to algebra achievement, the number of items where students show a misconception about variables is related to their performance on an algebra ability test that consisted of items from a standardized test. The finding that the misconceptions diagnosed on the current assessment are related to performance on standardized test items accentuates the need to address misconceptions.

The construct validity of the scores on the diagnostic test was also examined and supported by looking at its correspondence to students' transfer behavior on the cognitive transfer problems. Students transferred their correct and misconception responding at rates higher than chance for both near transfer problems (58% and 20% above chance, respectively) and far transfer problems (55% and 12% above chance, respectively). Meanwhile, incorrect responses were transferred at a rate below chance for both near and far transfer problems. The differences in these transfer percentages highlights the differences in the conceptual bases (or not) of these response kinds. Error-mistake responses are not thought to be conceptually (rule) based and hence would be less stable in reasoning and therefore less likely to transfer. Correct responses are thought to be conceptually rooted and thereby more stable in reasoning, so it was not surprising that correct responses transferred at the highest rates. The finding that misconception responses also transfer to near and far transfer problems at rates above chance suggests that misconceptions are also conceptually rooted.

A substantial portion of participants (28%) chose a misconception answer three or more times, suggesting that misconceptions about variables are quite common. This finding provides support for previous work that has shown the ubiquitous nature of variable misconceptions (Kuchemann, 1978; Stacey & Macgregor, 1997).

Findings also suggest that misconceptions are entrenched in this sizable group of students that tended to answer with misconception responses. These students rely on misconception reasoning when solving difficult problems as well as when solving easier problems. Hence, misconception reasoning appears robust and to guide the mathematical performance of these students regardless of a significant contextual factor, such as degree of problem difficulty.

The data also point to developmental trends in the formation of student misconceptions. Prior research has shown show that students form misconceptions early on in schooling, during the elementary school years, and enter algebra courses with misconceptions (McNeill & Alibali, 2005). The current results show that high school students had more misconceptions than middle school students suggesting that misconception reasoning increases from middle- to high-school. It is important to note that the data from the current study cannot determine the reason why misconception reasoning increases from middle- to high-school. Previous research suggests that the continued development of misconceptions may be linked to the teaching of algebra itself. For example, Macgregor and Stacey (1997) found that some curriculum materials present variables as abbreviated words, and the use of these curriculum materials coincided with the likelihood that students treat a variable as a label for an object; e.g. c could stand for ‘cat’, so 5c could mean 5 cats. The current finding that misconceptions become more common and entrenched from middle- to high-school is of particular concern, given the importance of learning algebra for students' future success (Achieve, 2007).

The findings from the present research underscore the importance of, and need for, a formative diagnostic assessment. The fact that the newly developed assessment is short and can be given with paper and pencil, enables teachers to quickly administer it to individual students or an entire class. Because the scores on the test exhibit strong reliability and validity, a teacher can assume with some confidence that when a student selects a misconception answer, he/she likely harbors that misconception. Specifically, a misconception answer on item(s) 1, 2, and/or 3 suggests an erroneous understanding that a variable can be ignored. A misconception answer on 4, 5, and/or 6 suggests an erroneous understanding that a variable is a label for an object. The erroneous interpretation that a variable is a specific unknown is indicated by a misconception response on problems 7, 8, and/or 9. Accordingly, the test results can enable teachers to engage in differentiated instruction for individual students.

In addition, the results reveal that some of these erroneous understandings are more common than others. We found that the belief that a variable is a label for an object was the most common misconception, followed by thinking a variable
is a specific unknown. The least common misconception was thinking a variable can be ignored. Yet even this misconception occurred in almost 15% of responses. These larger group patterns about the frequency with which different types of misconceptions occur is useful as teachers plan group activities and lessons.

Overcoming misconceptions is a matter of conceptual change, as opposed to conceptual growth. Conceptual change is more difficult to achieve than simple growth and requires specific instructional strategies. Use of this diagnostic test can inform teachers as to when these strategies are called for. These strategies include, but are not limited to, the use of model based reasoning (which helps students construct new representations that vary from their own intuitive theories), diverse instruction (when a few examples are presented that challenge multiple assumptions instead of many examples that challenge just one assumption), and engaging in interactive conceptual instruction (ICI). See the American Psychological Association (2011) website as listed in the reference section for further and more detailed discussion of these instructional strategies.

In addition, instructional strategies to achieve conceptual change in algebra specifically have been recommended. Welder (2011) encourages teachers to have students write out abbreviated words when introduced to variable (i.e. ‘7 feet’ instead of ‘7ft’) as well as literal translations. It is also recommended that students use language based techniques to first learn to interpret variables. For example, students should learn to read $6p=5$ as “6 times the number of professors equals the number of students”, even though this language based method may be more time-consuming (Gardella & Tong, 2002; Welder, 2011). Teachers can also explicitly talk about and pose specific questions that address particular misconceptions. See also the STAAR (Supporting the Transition from Arithmetic to Algebra Reasoning) program (Stephens, 2005; Welder, 2011) and the book Children’s Understanding of Mathematics: 11–16 (Hart, 1981). The current formative assessment can be used to identify students that may benefit the most from the strategies presented in these resources.

It is also important to note that the current diagnostic test is not limited to teacher use in the classroom. It can also serve as an instrument in research. For example, the diagnostic test can be used to categorize experimental subjects into students that do and do not hold variable misconceptions of certain types. Moreover, the diagnostic test and isomorph problem set can be used as pre-post test measures, respectively. Hence the diagnostic test is multi-functional, having both classroom and scientific uses.

Acknowledgements

The present research was supported by the Institute of Education Sciences, U.S. Department of Education through grant no. R305H040099. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. We thank Michael Russell and Rachel Kay for their valuable contributions.

Appendix. Diagnostic test items

This is also accessible for teachers to access for free at:
http://www.dartmouth.edu/~edpsychlab/edpsychlab/PUBLICATIONS.html

1) $n$ is a whole number greater than 0 and less than 5. How many values of $3n$ can there be?

A 0
B 3 (m)
C 4
D 5

2) $m$ is a positive whole number. How many possible values can $10m$ have?

A 5
B 10 (m)
C 20
D ‘Infinitely many’

3) Simplify $3m + 5 – 2m + 1$.

A $7m$
B 10
C $m + 6$
D $7m + 8$

4) In the expression $t + 4$, what does $t$ represent?

A 10
B 20
C time (m)
D ‘Any number’

5) At a university, there are six times as many students as professors. This fact is represented by the equation $S = 6P$. In this equation, what does the letter $S$ stand for?

A number of students
B professors
C students (m)
D none of the above
6) Latoya and Keith dropped a ball from various heights and measured the height of each of the bounces. They recorded their data in the chart below.

<table>
<thead>
<tr>
<th>Height from which ball was dropped (d)</th>
<th>40 in.</th>
<th>50 in.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height of each bounce (b)</td>
<td>20 in.</td>
<td>25 in.</td>
</tr>
</tbody>
</table>

Which equation best shows the relationship between the height from which the ball was dropped and the height of the ball’s bounce?

A: \( h_b = h_d + 20 \) (m)
B: \( b = 2d \)
C: \( b = d + 30 \)
D: \( b = \frac{1}{2}d' \)

7) How many different values can the expression \( k + 8 \) have if \( k \) can be replaced by any number?

A: One (m)
B: Infinitely many
C: Eighty
D: Zero

8) Trees are cut and new ones are planted. The data are shown below.

<table>
<thead>
<tr>
<th>Number of trees planted (p)</th>
<th>Number of trees cut (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

Which equation that will allow you to predict the number of trees planted (p) given the number of trees cut (c)?

A: \( c = 2p \)
B: \( c = p + 3 \) (m)
C: \( c = 4p \)
D: \( c = 2p + 100 \)

9) Rita put some hummingbird feeders in her backyard. The table shows the number of hummingbirds that Rita saw compared to the number of feeders. Bird-Watching

<table>
<thead>
<tr>
<th>Number of Feeders (f)</th>
<th>Number of Hummingbirds (h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

Which equation best describes the relationship between \( h \), the number of hummingbirds, and \( f \), the number of feeders?

A: \( h = 11f \)
B: \( h = 2f + 1 \)
C: \( h = f + 2 \) (m)
D: \( h = f + 6 \)

References


