Students' understanding of the concept of a variable

Elizabeth Warren, Australian Catholic University

This paper explores the use of written tests and semi-structured interviews in ascertaining students' understanding of the concept of a variable. A written algebra test was administered to 379 students and from the results students were selected for a semi-structured interview. The types of questions asked and the medium in which they were asked appeared to influence the responses given. It is conjectured that these issues must be considered when endeavouring to reach a richer understanding of the students' perception of the concept of a variable.

Algebra has long held a place of distinction in the mathematics curriculum (NCTM, 1989; National Statement for Australian Schools, 1991). Few have contested the importance of algebra as it is seen as 'the language through which most mathematics is communicated' (NCTM, 1989, p. 150). Critical to the algebraic domain is the variable construct. The literature identifies distinct stages students pass through when reaching an understanding of the concept of a variable.

Stages of understanding the concept of a variable

Through his large-scale study of students' interpretations of literal terms, Kuchemann (1978, 1981) identified six different stages. These were based on stages developed by Collis (1975) and are as follows:

Letter evaluated: Students assign numerical values to letters at the outset of a problem. For example, when asked to describe the expression $2+3x$ children often assign a value to x, such as 1, and compute the answer. Thus $2+3x=2+3\times1=5$.

Letter not used: Here, students ignore the letters, or at best acknowledge their existence but without giving them meaning. For example, the algebraic expression $2x+8y+3x$ is equated to $13xy$. Such an answer is obtained by simply adding up all the numbers, then writing down each letter that occurs.

Letters used as objects: Here, students regard the letter as shorthand for an object or as an object in its own right. For example, $2a+3b$ represents adding 2 apples to 3 bananas. This is referred to in the literature as 'fruit salad' algebra (Booth, 1988).

Letters used as a specific unknown or constant: Students perceive the letter as a specific but unknown number. For example, the expression L+M+N would never equal L+P+N as N cannot equal P. Even though both N and P are acknowledged as variables they must always be different values from each other as they are represented by different letters of the alphabet.

Letter used as a generalised number: Here, students perceive the letter as representing, or at least as being able to take, several values rather than just one. For example, if students are asked to list all the possible values for the expression $x+y=10$, they will list more than one of the whole numbers which will satisfy the condition. However, they tend not to realise that all numbers that satisfy the condition are required.

Letter used as a variable: In this final stage, students see letters as representing a range of unspecified values including all the rational and irrational numbers. Students also understand that the variable is defined by its relation to other terms in the expression. For example, in the expression $x+y=10$ the value of x is dependent upon the value of y.

Kuchemann (1981) found that most students (aged 13-15 years) could not cope consistently with items that required the use of the letter as a specific variable. Kuchemann (1981) stated that generally the first three stages indicated a low level of response.

Common errors and misconceptions of the concept of a variable

The first five stages delineated by Kuchemann (1981) also represent common misconceptions children experience in interpreting algebraic symbols. For example, when examining algebraic expressions, students may assign a numerical value to the variable, ignore the variable, view the variable as a specific unknown, or see the variable only consisting of whole numbers. In addition to the above misconceptions a further three are identified in the literature. These are:- first, changing the variable symbol as changing the referent. That is, different variables must take on different values. For example, the expression $3n$ is not the same concept as $3x$ as n and x could never be equal (Booth, 1984; Chalouh & Herscovics, 1983; Wagner & Parker, 1993). Second, assigning numerical values to letters according to their rank in the alphabet, for example, $a=1$ and $z=26$, or if $x=3$ then $y=4$ and $z=5$. (Booth, 1984; Chalouh & Herscovics, 1983). Third, assigning the letter as a subdivisional label, for example, 3a refers to the first part of the problem (Chalouh & Herscovics, 1983).

Most of these misconceptions were delineated from the results of large studies involving written tests. The types of questions asked and the medium in which they were asked could influence the responses given. Through written responses students may not be able to exhibit their full understanding or misunderstanding of the concept of a variable. This paper compares the results from written responses and a semi-structured interview and explores how each contributes to reaching a richer understanding of students' perceptions of the concept of a variable.

Methodology

This study comprised whole class testing to select students for in-depth interviews. The sample for the whole class testing comprised 379 students from one independent coeducational school and one state coeducational school in the metropolitan area. The students' ages ranged between 12 years and 2 months and 15 years and 10 months. Both schools chosen for the study consisted of students from lower-middle socio-economic status, with a variety of ethnic backgrounds represented. Both schools were large metropolitan schools comprising up to two thousand students.

Whole class test: The whole class test was administered as a written test and completed within a one-hour period. The test consisted of three components, namely, generalising from visual patterns, generalising from tables of data, and understanding the concept of a variable. The results from 3 of the 8 questions selected for the variable component of the written test are discussed in this paper.

Semi-structured interview: The results for each component were ranked and three students were selected at the 1st, 25th, 75th and 100th percentile for each of the written components. These particular percentiles were chosen to highlight the difference in levels of understanding of the concept or a variable. Each interview was videotaped and transcribed.

Instrument

From the results of the written test 2 questions were selected for inclusion in the semi-structured interview, and a third question was adapted for the semi-structured interview. These 3 questions were believed to probe students' view of the variable, their need for closure of algebraic expressions, their propensity to concatenate algebraic expressions, and why students allowed "= constant" to limit possible values for the variable. The selected questions were:

Question 1. **This question is about t+t and t+4.**

(a) **Is t+t ever larger than t+4? If so when?**

(b) **Is t+4 ever larger than t+t? If so when?**

(c) Are $t+t$ and $t+4$ ever equal? If so when? (Modified Harper, 1979)

This item was modeled on an item used by Harper (1979) and Quinlan (1992). The item used in their studies was as follows:

This is a question about t+t and t+4

(a)	*Which is larger, t+t or t+4? WHY?*	*(b)*	*When is t+t larger?*
(c)	*When is t+4 larger?*	*(d)*	*When are they equal?*

In the interview stage of the pilot study, concern was expressed about the inconsistency in the wording of the Harper item. For example, *"When is t+t larger?"* seemed to alert to the fact that perhaps $t+t$ could be larger. This proved confusing for students who had chosen $t+4$ for part (a). Hence the question was reworded so that such inconsistencies no longer existed.

Question 2. **For a school excursion, 3 buses take f students each and 4 cars take g student each.**

(a) Give the total number of students taken by these buses and cars.

(b) One car leaves early with g students. How many students remain? (Quinlan, 1992b)

This question was created to obtain information on students' ability to interpret the meaning of letters when they referred to a real-life context and to carry out operations on numerical variables without knowing their values.

Quest 3. (a) If $c+d=10$, tick ALL the meanings that c could have:

3 10 12 7.4 the number of apples in box

an object like a cabbage an object like an orange

(b) If $c+d=10$, what happens to d as c gets bigger?

(c) If $c+d=10$, and c is always less than d, what values may c have(Harper, 1979,Kuchemann, 1980 & Quinlan, 1992b

The origin of this item is threefold. It is based on the CSMS project question (Hart, 1981) *"What can you say about c if c+d=10 and c is less than d,"* (Kuchemann 1980, p. 67) and on Harper's (1979) equation task, *"If x+y=10 when is x less than y."* Kuchemann (1980) regarded this question as one which required students to view letters as generalised numbers. Quinlan (1992, p. 106) added parts (a) and (b) to the original question. Part (a), measured the types of possible meanings students are prepared to accept for alphabetic symbols in algebra. Part (b) tested students' understanding of the covarying relationship between the two variables (i.e., as the value of c varies so does the value of d). In the semi-structured interview students were asked *"If p+m=12, put a ring around all the possible values p can have. 4, 12, 15, 0, 3.9, -2, an object like a pear, the number of children in the class."* Students were then asked to articulate their reasons for including or excluding certain values.

Results

Written component

Question 1.: For each part of question 1, students were required to give two responses. Students were firstly asked to reply either in the affirmative or negative to the question and then were asked to elaborate on their response with an appropriate reason. Table 1 summarises the results for this question.

Table 1

Percentage of Students who correctly responded to Each Part of Question 5

Question 1	Correct response	Valid reason
(a) Is t+t ever larger than t+4? If so when?	46.4	28.8
(b) Is t+4 ever larger than t+t? If so when?	45.9	29.0
(c) Are t+t and t+4 ever equal? If so when?	43.0	36.1

Less than half the students answered this question. The results seem to indicate that giving a valid reason for the equality situation was easier than the other two situations. The rewording of the Harper Quinlan question appeared to result in a greater percentage of students reaching correct solutions. While Quinlan (1992) found that 50% of his sample (n=517) responded correctly, 47% (n=241) of those who offered a correct response had completed at least 2 more years of schooling with 21% being in their final year. The students' responses to question 1 were also assigned a level according to their understanding of the relationship between $t+t$ and $t+4$. Each level represents a developing understanding of the concept of a variable. Table 2 describes each level with the percentage of students whose responses were allocated to that particular level.

Table 2

Levels of Responses for Question 1

Level	Description	% response
0	Did not answer	56.2
1	Letters represented objects	0.5
2	Focus on the constant rather than the variable	4.2
3	Use of only one trial example to reach a solution	4.7
4	Correct response for specific values	5.0
5	Correct response in general terms	29.3

The majority of students did not respond to this question. It appeared that only 29.3% of the students saw the variable in general terms.

Question 2: Responses to both part (a) and part (b) seemed to fall into four distinct levels Table 3 presents a summary of a description of the levels together with the percentage of students whose responses were considered to be at that level.

Table 3

Levels of responses to Question 2

Level	Description	% response	
		2 (a)	2(b)
0	Did not attempt or assigned a number	29.0	31.4
1	Incorrect algebraic response	14.5	21.4
2	Correct response but the need for closure	7.7	10.8
3	Correct response	48.5	36.4

Students found part (a) easier to respond to than part (b). This is not surprising given that a correct response to (b) appears to be contingent on a correct response to (a). Also the need for closure was more prevalent for (b).

Question 3: Table 4 summarises the percentage of students who chose the various options in part (a) of question 3.

Table 4

Percentage of Responses to each Choice for Question 3(a)

	Choice						
	3	10	12	7.4	number in box	object cabbage	object orange
accepted	76.5	51.5	11.9	66.5	40.4	22.7	21.4

The majority of students accepted 3 and 7.4 as possible values for the variable c but were less confident in allowing c to be 10. When c is 10 the value of d is zero. Only eleven percent accepted c as 12. When c is 12 the value of d is negative. These findings seem to indicate that the nature of the product, "equal ten," of the equation plays a significant role in limiting the permissible values for the associated variables. This trend was difficult to delineate in the Kuchemann (1980) format of the question. For the majority of students, it seems that '=10' represents a barrier beyond which values cannot be selected. This could reflect reluctance by the students to assign zero or negative values to the variable *d*. This issue was investigated in the semi-structured interview. Only 40.4 percent of the students accepted that *c* could be the number of apples in a box, whereas up to 22.7 percent allowed *c* to stand for an object such as an orange or cabbage. This seems to indicate that the students had difficulty in accepting the variable as a generalised number and in differentiating between this interpretation and the variable as an object.

Levels were also developed for classifying responses to part (b) and part (c). For part (b), the levels indicated whether the students could see the covarying relationship between *c* and *d* in the equation $c+d=10$. At the lowest level students failed to see the interrelationship between *c* and *d*. For part (c), the levels indicated the possible values students would accept for *c* if *c* is always less than *d* in the equation $c+d=10$. A level 4 response was '<5' and a level 5 response was '<5' but also included 0, negatives and fractions. This distinction was made as it was felt that some student's response of '<5' might only consist of the positive integers less than 5. The following tables describe the levels and record the percentage of students whose responses were classified at that particular level.

Table 5

Levels of Responses for Question 3(b) "If c+d=10, what happens to d as c gets bigger"?

Level	Description	% response
0	Did not respond	20.1
1	No change (Fails to see the interdependence between the 2 variables)	6.3
2	Bigger (sees a relationship between the 2 variables; incorrect response)	5.3
3	Smaller (i.e., as c gets bigger d gets smaller)	68.3

Table 5 indicates that 68.3% of the students reasoned correctly that *d* would get smaller.

Table 6

Levels of Responses for Question 3(c) "If c+d=10, and c is always less than d, what values may c have?"

Level	Description	% response
0	omitted response or incorrect list of numbers e.g., 1,2,3,4,5,..10	29.4
1	one or two numbers as response e.g., 3,4	4.5
2	1,2,3,4 – included only integer responses	42.9
3	included 0, negatives and/or fractions (4 & under or 4.99 down to 0)	20.6
4	< 5	2.1
5	< 5 and included 0, negatives & fractions	0.05

As can be seen from the above results very few students responded correctly to question 3(c). Over half of the students tested gave at least four values for *c*. Kuchemann (1981, p. 108) found that nearly 40% of his sample of 14 year old students (n=1000) gave only one value for *c* and that 30% gave four or more. For this study, approximately 43% of the students gave four values for *c*. As indicated by Quinlan (1992), it appears that the inclusion of the first two components helps students frame appropriate answers to (c).

In summary these 3 questions from the written test indicated that when examining the relationship between two variables in an expression of the form $x+y=c$, where c is a constant, the value of the constant seems to play an important role in the permissible values students will accept for the variables themselves. It appears that the constant, c, represents a barrier. Students exhibited a reluctance to choose values for either x or y that were greater than c. Even students who were considered by their teacher to be very capable of understanding algebraic concepts, included the letter standing for an object in a number of their responses. A very small percentage of the students perceived the variable as a generalised number. Most students tended to see the variable as being represented by up to 4 specific integer values. These trends were investigated further in the semi-structured interview.

Semi-structured interview

Twelve students were chosen to be interviewed on their understanding of the concept of a variable. From the literature it seems that students' perceptions of the concept of a variable range from the letter standing for an object through to the variable as a generalised number. The semi-structured interview attempted to explore some of these understandings. The following section presents a summary of students' perceptions.

Variable as an object: The results for question 3 from the written algebra test indicated that 22.7% of students allowed c to stand for 'an object like a cabbage' and 21.4% allowed c to stand for 'an object like a orange' (see Table 4). From an analysis of the written responses, some of the more able students seemed to hold this misconception. The inclusion of "object like a pear" in the third interview question probed why students accepted the variable as standing for an object. Not one of the twelve students interviewed accepted p as "shorthand for pear." When questioned, those students who ringed 'an object like a pear' seemed to assign a number to p. For example, Jason at the 1st percentile said, "Yes it could be about the number of pears and number of apples." Michelle, also at the 1st percentile said, "Could be 14 pears." One student at the 100th percentile could clearly articulate why 'an object like a pear' was not a valid option. Mary said, "Object like a pear and 12 have no relationship," indicating that she viewed 'pear' as an object and not as a number. The two other students at the 100th percentile said, "p could be an object like a pear only if m was 11 pears and then you would have 12 pears." The interview seemed to indicate that, contrary to the belief held when analysing the written component, simply ringing 'object like a pear' did not necessarily mean students saw p in expressions like $p+m$ standing for pear, but rather as 'a pear', that is, 'one pear'. When questioned, students who ringed this option, expressed valid reasons for its inclusion in their options. This seems to indicate that relying solely on written responses could lead to erroneous conclusions with respect to students' understanding of the concept of a variable.

Specific unknowns: From the students' responses there seemed to be a growth of acceptable values for the variable. Students at the 1st and 25th percentile made decisions about the two expressions $t+t$ and $t+4$ using one specific value for the variable. Adam said, "$t+t$ is bigger when $t=5$, $t+4$ is bigger when $t=2$ and they are the same when t equals 4." By contrast, one student at the 75th percentile reached a conclusion by using a series of specific unknowns for t. For example, Jenny said, "$t+t$ is bigger when t is 5 or 6. $t+4$ is bigger when t is 1,2, or 3." For this problem, the other two students at the 75th percentile perceived the variable as a generalised number, that is, "$t+t$ is bigger when $t > 4$ and $t+4$ is bigger when $t < 4$".

Generalised number: For the expressions $t+t$ and $t+4$, two students at the 75th percentile and three students at the 100th percentile immediately articulated the general case, that is, "$t+t$ is larger when t is greater than 4, $t+4$ is larger when t is less than 4 and they are equal when $t=4$." It was conjectured that they had an understanding of the variable as a generalised number, that is, thinking at level 5 (see Table 2). Students at the 1st and 25th percentiles either could not solve the problem nor needed specific values in order to reach a solution.

Closure: For question 2, 3 students at the 1st and 25th percentile wrote $c+f=$, indicating a need for closure. All the students at the 75th and 100th percentile did not need to close the expression, writing $c+f$.
Product component of an equation: For the expressions $p+m=12$ it seemed that the '=12' component of the question was not the only factor that limited the number of acceptable values for p. The addition sign before m also seemed to play an important role. The six students at the 1st and 25th percentile failed to accept 15 for the value of p. A common response was "15 plus anything is over 12," indicating that the value of m must be positive as it was preceded by '+'. Of particular interest was the fact that they all accepted '-2' as a value for p. One student stated that, "-2 was alright as it is below 12."The other six students at the 75th and 100th percentile accepted 15 as a valid value for p, stating that m would need to be negative.
Comparing t+t with t+4 by considering the constant only: Only two students seemed to compare $t+t$ with $t+4$ by focussing on the value of the constant. Lachlan, at the 25th percentile said, "$t+4$ would always be bigger because the 4 made it bigger." Lindsay, at the 100th percentile, also seemed to allow the value of the constant to influence his decisions. He said, "for $t+4$ you add 4 to a variable whereas $t+t$ has two variables so it must always be bigger," suggesting that an expression containing two variables will always be bigger than an expression only containing one variable. Adam, at the 25th percentile, continually asked if the t's were all the same. He seemed to want to substitute different values for t in the two expressions.
Concatenation of algebraic expressions: Three students exhibited a need to concatenate the variables. All of these students were either at the 1st or 25th percentile. For question 1, Chris at the 1st and James at the 25th percentile indicated that they would have '*cf* trucks'. When comparing t+t and $t+4$, Chris also suggested that, "$t+t$ is $2t$ and $t+4$ is $4t$ so $t+4$ would be bigger."
In summary the results of the interviews indicated:-Students at the 1st and 25th percentile were characterised by their need to evaluate, close and concatenate algebraic expressions, and by perceiving the variable as representing a specific unknown. Two students at the 75th percentile and the three students at the 100th percentile perceived the variable as a generalised number and did not concatenate or close any of the algebraic expressions.

Discussion and conclusion

Both the results of the written test and the interview support Kucheman's stages of understanding of the concept of a variable. Not only did the less able students, students at the 1st and 25th percentile, seemed to perceive the variable as representing a specific unknown but they also exhibited a need to close algebraic expressions. By contrast, the more able students did not need to close algebraic expressions and they perceived the variable as a generalised number.

The results also seem to indicate that care must be taken when conjecturing about students understanding from written responses. First, the framing of the question itself could limit the types of responses given. The use of appropriate prompts could help elicit students' full understanding of the concept of a variable. For example, as indicated by Question 3 in the written test, the inclusion of (a) and (b) helped students frame a response to (c), and for question 1 the rewording resulted in a greater percentage of students reaching correct solutions.

Second, the medium in which questions are posed needs also to be considered when reaching conclusions from students' responses. From an analysis of the results of the written algebra test it seemed that many students perceived the variable as a letter standing for an object. The results of the interview seemed to indicate that this assumption was too simplistic. None of the students interviewed believed that p was "shorthand for pear." All the students who selected p as standing for an object like a pear, had valid reasons for its inclusion and these reasons relied on p standing for a number of pears and not pear. Also, for the expressions $p+m=12$ the product component (=12) is not the only factor that limited students' thinking. It seemed that the placement of '+' before the variable also played a key role. Students at the 1st and 25th percentile believed that '$+m$' meant that the value of m must always be positive. This seems to indicate that some students are not capable of delineating between '+' as the operation of addition and '+' indicating a positive value. Both of these misunderstandings are difficult to identify in a written test format.

References:

Booth, L. (1984). *Algebra" Children's strategies and errors. A report of the strategies and errors in secondary mathematics project.* Windsor: The Nfer-Nelson Publishing Company. Ltd.

Booth, L, (1988). Children's difficulties in beginning algebra' In A. F. Coxford & W. P. Shulte (Eds.), *The ideas of algebra K-12* (pp. 20-32). Reston, VA: NCTM, Inc>

Chalouh, L, & Hersovics N. (1983). The problem of concatenation in early algebra. *Proceedings of the 5th annual meeting of the North American Chapter of the international group PME:* Montreal

Collis, K. (1975). *The development of formal reasoning*. Newcastle: University of Newcastle.

Kuchemann D. (1978). Children's understanding of numerical variables. *Mathematics in School, 7*(4), 23-26.

Kuchemann D. (1984). Algebra. In K. Hart (Ed.), *Children's understanding of mathematics* (pp. 11-16). London: John Murray.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA. NCTM.

Quinlan, C. (1992). *Developing an understanding of algebraic symbols.* Unpublished doctoral dissertation, University of Tasmania, Hobart.

Wagner, & Parker, (1993). Advancing algebra. In P. S. Wilson (Ed.), *Research ideas for the classroom: High school mathematics* (pp. 120-139). New York: Macmillan Publishing Company.