Visual Salience of Algebraic Transformations[1]

David Kirshner, Louisiana State University

Thomas Awtry, Louisiana State University, Alexandria

Journal for Research in Mathematics Education, 35(4), 224-257

July 2004

Key words: Algebra; Cognitive theory; Epistemology; Language and mathematics; Logic and proof; Oral communication; Reform in mathematics education; Spatial visualization

Contact Information:

David Kirshner, Mathematics Education, 103 Peabody Hall, Department of Curriculum & Instruction, Louisiana State University, Baton Rouge, LA 70803-4728; dkirsh@lsu.edu

Thomas Awtry, Mathematics, Louisiana State University at Alexandria, Department of Mathematics and Physical Sciences, 8100 Hwy 71 South, Alexandria, LA 71302; tawtry@lsua.edu

[1]We express appreciation to John St. Julien, Ken Koedinger, Paul Ernest, and James J. Madden for help with technical arguments in philosophy, connectionist theory, and mathematics, and to two editors and numerous anonymous reviewers of *JRME* who have contributed substantially to the shaping of this article.

Visual Salience of Algebraic Transformations

Abstract

Information processing researchers have assumed that algebra symbol skills depend on mastery of the abstract rules presented in the curriculum (Matz, 1980; Sleeman, 1986). Thus, students' ubiquitous algebra errors have been taken as indicating the need to embed algebra in rich contextual settings (Kaput, 1995; National Council of Teachers of Mathematics [NCTM], Algebra Working Group, 1998). This study explored a non-representational account of symbolic algebra skills as feature correlation within the visual field. We present evidence that algebra students respond spontaneously to the visual patterns of the notational display apart from engagement with the declarative content of the rules. Thus, persistent algebra errors may reflect disengagement from declarative content rather than inability to deal with it. We sketch a Lexical Support System designed to sustain students' engagement with the declarative content of algebraic rules and processes, thus complementing the exciting curricular possibilities being developed for referentially rich algebra.

> The Study of Algebra may be pursued in three very different schools, the Practical, the Philological, or the Theoretical, according as Algebra itself is accounted an Instrument, or a Language, or a Contemplation; according as ease of operation, or symmetry of expression, or clearness of thought, (the *agere*, the *fari* or the *sapere*,) is eminently prized and sought for. ...The felt imperfections of Algebra are of three answering kinds. ...The Philological Algebraist complains of imperfection, when his Language presents him with an Anomaly; when he finds an Exception disturb the simplicity of his Notation, or the symmetrical structure of his Syntax; when a Formula must be written with precaution, and a Symbolism is not universal.[2]
>
> (Sir William Rowan Hamilton, 1837, p. 293)

That application and theory each has motivated mathematical engagement is a commonplace observation. However, the philological impulse Hamilton (1837) attributed to the pursuit of algebra is less familiar. Typically we regard the notations and symbols of mathematics as conferring "a power to name and rename, to transform names, to use names and descriptions to conjure, communicate and control our images, our mental worlds" (Pimm, 1995, p. 1). In other words, it is the images that populate our mental worlds that are the focus of mathematical inquiry; notations and symbol systems are a gateway to these images, not a motivational focus in their own right. What impulse to algebra might Hamilton have been alluding to with such phrases as "symmetrical structure of his Syntax?"

[2]We are grateful to Roberta Mura for bringing this quotation to the attention of the first author.

The historical context may provide some clues. In Hamilton's time, the use of variable symbols was propelling algebra well beyond any theoretical grounding then available. As Kline (1980) observed,

> In the first half of the 19th century, the logical foundation of algebra was also noticeable by its absence. The problem in this area was that letters were used to represent all types of numbers and were manipulated as though they possessed all the familiar and intuitively acceptable properties of the positive integers. ... It seemed as though the algebra of literal expressions contained a logic of its own (p. 158).

Hamilton's contemporary, George Peacock (1833), attempted to deal with this problem by fiat. In his Principle of the Permanence of Equivalent Forms, he "dogmatically affirmed" (Kline, 1980, p. 159) the universality of properties established for whole numbers: "Whatever algebraical forms are equivalent when the symbols are general in form but specific in value [positive integers], will be equivalent likewise when the symbols are general in value as well as in form" (Peacock, 1833, quoted in Kline, 1980, p. 159).

While some contemporary mathematicians like Gregory, De Morgan, and Hankel adopted Peacock's strategy (Kline, 1980), Hamilton was highly agitated by the absence of theoretical foundations. In fact he found mathematics to be so "unsatisfactory as an exercise of the mind" (Cajori, 1924, p. 304) that he railed famously against its incorporation into general education: "If we consult reason, experience and the common testimony of ancient and modern times, none of our intellectual studies tend to cultivate a smaller number of the faculties, in a more partial manner, than mathematics" (Hamilton, 1836, quoted in Cajori,1924, p. 304). Perhaps, then, it was a keen awareness of the limitations of algebra as a theoretical pursuit that led Hamilton to appreciate an aesthetics of form as intrinsic to algebra's fascinations.

Rather than pursue such historical speculations, we set out, here, to examine an aesthetics of algebraic form psychologically. In particular, we explore students' initial response to algebra rules, and present evidence that from the very start they are receptive to the visual structure of such rules separate and apart from intellectual engagement with the declarative content. These results are at odds with the usual cognitivist assumption that human intellectual skills rest on acquiring or developing well structured algorithms or rules—a matter we take up in the Theoretical Frameworks section of this article. This leads us to a critique of the current algebra reform movement with its insistence that rich contextual settings for algebra are necessary to "provide the substance from which and about which to reason" (NCTM Algebra Working Group, 1998, p. 164). We take a different lesson from the failures of the traditional algebra curriculum. Rather than conclude that students' ubiquitous symbol manipulation errors reflect failure in dealing with algebra as abstract, decontextualized rules, we see an intrinsic aesthetics of visual form as systematically drawing students away from intellectual engagement. As a new curricular innovation, we outline a Lexical Support System designed to sustain students' engagement with the declarative content of algebraic rules and processes, an agenda we see as complementary to the burgeoning and exciting curricular possibilities being developed for referentially rich algebra. In the final section, we return to the ontological status of elementary algebra as we distinguish our interests in algebra as mathematical *method* from the New Math era of the 1960s in which abstract algebra *content* was incorporated into the school curriculum (e.g., Haag, 1961).

THEORETICAL FRAMEWORKS

The traditional school algebra experience produces a strong tendency for students to generate patterns of incorrect expression transformation consistent with applying "mal-rules" (Sleeman, 1986) like those indicated in the left column of Table 1. Such response patterns have been widely observed, documented, and classified for a quarter of a century (Booth, 1984; Bundy & Welham, 1981; Carry, Lewis, & Bernard, 1980; Davis, 1979; Greeno, 1982; Matz, 1980; Payne & Squibb, 1990; Radatz, 1979; Sleeman, 1982, 1984, 1986; Wagner, Rachlin, & Jensen, 1984).

Table I

Mal-rules and Correct Rules

Mal-rules	Correct Rules
$(a + b)^c = a^c + b^c$	$(a + b)c = ac + bc$
$\sqrt[c]{a + b} = \sqrt[c]{a} + \sqrt[c]{b}$	$\sqrt[c]{ab} = \sqrt[c]{a}\,\sqrt[c]{b}$
$a^{mn} = a^m a^n$	$a^{m+n} = a^m a^n$
$a^{m+n} = a^m + a^n$	$a(m + n) = am + an$
$\frac{a}{b + c} = \frac{a}{b} + \frac{a}{c}$	$\frac{b + c}{a} = \frac{b}{a} + \frac{c}{a}$
$\frac{a + x}{b + x} = \frac{a}{b}$	$\frac{ax}{bx} = \frac{a}{b}$

What is so confounding about these errors is their superficial character. Rather than reflecting mis*understanding* of the *meaning* of correct algebra rules, they seem to indicate nothing more substantial than mis*perception* of the *forms* of the correct rules given in the right column of Table 1. For instance, Thompson (1989) spoke of algebra students as "prone to pushing symbols without engaging their brains" (p. 138). Similarly, in his landmark study Erlwanger (1973) observed:

> One may be tempted to treat this kind of talk as evidence of an algebraic concept of commutativity. But, in view of the whole picture of Benny's concept of rules, it appears more likely that it involves less awareness of algebraic operations than it does awareness of patterns on the printed page. (Note to p. 19)

Classical Artificial Intelligence Models

Ignoring this intuition that percept, rather than concept, often governs students' algebraic symbol manipulation, virtually all previous theorizations of algebraic skill development have regarded conscious, declarative knowledge as foundational. This is because the classical artificial intelligence (AI) paradigm within which such skills have been studied is inherently *representational* (Brooks, 1991; Clancey, 1999; Dupuy, 2000). Classical AI represents cognition as a serial process in which tokens, or instances of symbols, are manipulated according to fixed rules (Newell & Simon, 1985). The mode of representation is direct. When a model is interpreted, rules incorporated into the model reflect critical relationships among the elements (Haugland, 1985).

Within that tradition, theorists have solved the problem of accounting for students' acquisition of algebraic rules in the obvious way: as issuing from the explicit declarative content of the algebra curriculum. Thus Carry, Lewis, and Bernard (1980) speak of students coming to know "the legal moves of the algebra game" (p. 2). More explicitly, Matz's (1980) proposal

> idealizes an individual's problem-solving behavior as a process employing two components. The first component, the knowledge presumed to precede a new problem, usually takes the form of a rule a student has extracted from a prototype or gotten directly from a textbook. For the most part these are basic rules (such as the distributive law, the cancellation rule, the procedure for solving factorable polynomials using the zero product principle) that form the core of the conventional textbook content of algebra. (p. 95)

Matz (1980) explored how error patterns arise through the (mis)application of "*extrapolation techniques* that specify ways to bridge the gap between known rules and unfamiliar problems" (p. 95). In other words, mal-rules (such as those given in the leftmost column in Table 1) are overgeneralizations of the correct rules gained as explicit, declarative knowledge from the curriculum.

Such models do not negate the possibility that visual patterning may eventually come to dominate the cognitive processes associated with performance of algebra tasks. For instance, Davis (1984) spoke of what he called VMS sequences (visually moderated sequences) which “can be thought of as a visual cue V_1 which elicits a procedure P_1 whose execution produces a new visual cue V_2, which elicits a procedure P_2,... and so on” (p. 35) until the problem is completed. Indeed, concepts and percepts must be conceived not as fully independent of one another, but as supporting and reinforcing one another in an iterative process (Rittle-Johnson, Siegler, & Alibali, 2001). But such classical AI theories of skill development invariably posit the precedence, and hence the primacy, of declarative knowledge over visual patterning.

Perhaps the best known and most detailed theory of skill acquisition is the ACT theory of John Anderson and his colleagues. Through its many iterations (ACTE, ACT*, and ACT-R), ACT has maintained a fundamental distinction between “*declarative knowledge* [that] corresponds to things we are aware we know and can usually describe to others ... [and] *procedural knowledge* ... that we display in our behavior but that we are not conscious of” (Anderson & Lebiere, 1998a, p. 5). In these theories, the former always produces the latter: “Both in ACT* and ACT-R, new production rules ultimately derive from declarative knowledge” (Anderson & Lebiere, 1998b, p. 109). Thus the extant theories of algebra symbol manipulation, conceived within the classical AI tradition, take students’ explicit, declarative knowledge as the starting point of learning. It is this assumption that justifies the interpretation of persistent and widespread errors of symbol manipulation as evidence of students’ difficulties in apprehending algebra as an abstract domain of explicit rules.

Non-representational Models

In presenting evidence that students are immediately responsive to the visual structure of algebra rules, we contribute to a growing chorus of critique of the mentalist assumptions of classical AI. The persistent theme of this critique is that classical AI mistakes products of cognition—regularized, rule-like performance—for processes of cognition—rules in the head (Brooks, 1991; Clancey, 1999; Dupuy, 2000; Estep, 2003). As Clancey, 1999, expressed it,

> Traditional [AI] models serve [only] as a kind of *specification* for producing a scientific *explanation* of how the brain physically works. In short, the map is not the territory–pattern descriptions are not literally stored in the brain like recipes and conventional computer programs. Taxonomic models of knowledge roughly characterize what people know, but inadequately represent how categories form and how they relate structurally and temporally to constitute a conceptual *system*. (p. xiv)

The philosophical concern underlying these criticisms is that AI's assumption of explicit representations stored in the brain downplays the richness and immediacy of the cognitive connection to the world:

> According to Merleau-Ponty, in absorbed, skillful coping, I don't need a mental representation of my goal. Rather, acting is experienced as a steady flow of skillful activity in response to one's sense of the situation. Part of that experience is a sense that when one's situation deviates from some optimal body-environment relationship, one's activity takes one closer to that optimum and thereby relieves the "tension" of the deviation. ... As Merleau-Ponty [1962, p. 153] puts it: "Whether a system of motor or perceptual powers, our body is not an object for an 'I think', it is a grouping of lived-through meanings which moves towards it equilibrium" (Dreyfus, 2002, p. 378)

Within cognitive science connectionist modeling often is regarded as providing a non-representational alternative to the serial processing of traditional AI. Parallel distributed connectionist models provide a more neurologically plausible implementation of cognition (Freeman, 1991; Hobson, 1988), one that more easily and flexibly models basic cognitive functions like pattern recognition and associative memory retrieval which are not inherently well structured

(Bereiter, 1991; Lloyd, 1989; Rumelhart, Hinton, & Williams, 1986; St. Julien, 1994). For, whereas, "symbolic-processing models have had some success modeling performance under multiple constraints ... these models are cumbersome and brittle; they tend to break down when the stimulus conditions are poorly specified. Connectionist models are well suited for just such situations" (Haberlandt, 1997, p. 159).[3] Thus there is a credible body of empirical and theoretical work to support a departure from usual mentalist assumptions about algebra learning in order to explore the heretical notion that algebraic skills take root, not just from the declarative content of rules presented explicitly in the curriculum, but also from their visual patterns on the printed page.

However, we should caution that shifting from a representational to a non-representational view of knowledge entails a revision of what counts as a psychological theory of learning and performance. Our purpose in this study is not to operationalize a connectionist implementation of visual salience—merely to document the effects of visual salience on learning. In this endeavor, we characterize visual salience as an aesthetic sense of form of the sort Hamilton (1837) might have intended as part of the philological motivation for the pursuit of algebra. Such a notion of visual aesthetic is necessarily vague and indefinite in comparison with the hard edged and definitive models offered in the classical AI tradition that presumes knowledge is expressible as explicit rules.

Visual Salience in Algebra

Tables 2 and 3 present rules with greater and lesser visual salience, respectively. Run your eye from left to right across the rules presented in Table 2, and again for those in Table 3.

---

[3]Connectionist architectures are non-representational in the sense that they dispense with structured representations of AI that combine with one another according to syntactic rules. However, they are often referred to as representational in the looser sense that "they model the mental states that refer to and make sense of the world" (Bechtel & Abrahamsen, 2002, p. 156). Advocates of traditional AI have argued that connectionism must fail as a cognitive modeling architecture precisely because it lacks syntactically structured representations (see Bechtel & Abrahamsen, 2002, pp. 156-199, for a review of these criticisms and the counterarguments of connectionist theorists).

Table 2

Visually-salient Rules

$$x(y+z) = xy + xz$$

$$(xy)^z = x^z y^z$$

$$\sqrt[n]{xy} = \sqrt[n]{x}\sqrt[n]{y}$$

$$(x^y)^z = x^{yz}$$

$$x^{y+z} = x^y x^z$$

$$\frac{w}{x}\frac{y}{z} = \frac{wy}{xz}$$

$$\frac{xy}{xz} = \frac{y}{z}$$

$$\frac{x+y}{z} = \frac{x}{z} + \frac{y}{z}$$

Table 3

Non Visually-salient Rules

$$x^2 - y^2 = (x - y)(x + y)$$

$$\frac{w}{x} \div \frac{y}{z} = \frac{wz}{xy}$$

$$(x + y)^2 = x^2 + 2xy + y^2$$

$$x^3 + y^3 = (x + y)(x^2 - xy + y^2)$$

$$\frac{w}{x} + \frac{y}{z} = \frac{wz + xy}{xz}$$

$$x^0 = 1$$

The quality of visual salience is easy to recognize but difficult to define. Visually salient rules have a visual coherence that makes the left- and right-hand sides of the equations appear naturally related to one another. Elsewhere, we proposed two characteristics contributing to this visual coherence: repetition of elements across the equal sign; and a visual reparsing of elements across the equal sign (Awtry & Kirshner, 1994). Visual reparsing manifests itself as a dynamic visual displacement of elements. For example, boundary tensions created by the parse on the left-hand side of the equation may be resolved on the right-hand side. We compare this effect to an animation sequence in which distinct visual frames are perceived as ongoing instances of a single scene. Hence, we see the immediate connection between right- and left-hand sides as stemming from a sense that a single entity is being perceived as transformed over time. This relieves the observer of any obligation they might otherwise feel to articulate conscious connections between two separate entities.

Consider, for example, these two rules, both used in this study:

A) $(\frac{x}{y})(\frac{w}{z}) = \frac{xw}{yz}$, and

B) $(x - y) + (w - z) = (x + w) - (y + z)$

In Rule A, one easily perceives the right hand expression as resulting from physically adjoining adjacent elements of the left hand expression. In Rule B, one, likewise, could perceive the transformations in purely visual/spatial terms: the middle terms exchange location, as the middle sign exchanges with the extreme signs. But in this case, the complexity of the spatial operations creates a greater likelihood students would represent the two sides of the equation independently of one another as static entities, attending to the structure and composition of each. Thus we classify Rule A as having greater visual salience than Rule B, but recognize that visual salience/non visual-salience occurs on a continuum rather than as a dichotomy (see also Awtry, 1993, which reports on a pilot version of the current study). Note, also, that the perceived simplicity of Rule A relative to Rule B is not a reflection of the greater complexity of Rule B in a declarative sense. These two equations are structurally identical, with multiplication and division operations in Rule A systematically replaced with addition and subtraction operations in Rule B.

This analysis points toward a new explanation of students' persistent errors in algebra symbol manipulation. Take a close look at Table 1. Note that virtually all of the patterns of error are spawned from correct rules that are visually salient. For example, whereas students regularly overgeneralize the visually-salient rule $(xy)^2 = x^2y^2$ as $(x + y)^2 = x^2 + y^2$, they virtually never overgeneralize a non-visually-salient rule like $x^2 - y^2 = (x - y)(x + y)$ as, say, $x^2 + y^2 = (x + y)(x - y)$. This raises the possibility that such errors may stem, in part, from the visual salience of a rule, rather than just from it's declarative complexity–a matter we consider more fully in the Discussion section. In this respect, persistent errors may reflect lack of attention to the declarative content of algebra rules, rather than difficulty in processing such content.

The methodological point is worth stressing here. When pointed out, visual salience may seem an obvious phenomenon, in keeping with our intuitions about algebraic skill development (Erlwanger, 1973; Thompson, 1989). But by the usual standards of classical AI, our characterization of visual salience is unacceptably vague: It is not, and cannot be, captured in well-defined and definitive algorithms. However, adopting a non-representational view of cognition gives us an entree to explore such a vague construct as visual salience, and to pursue its implications for curriculum and instruction.

METHOD

This study used a two treatment experimental design to assess the cognitive basis of algebraic symbol manipulation skill. To guard against the reasonable assumption that skill development might progress from declarative origins to non-conscious proceduralization (Anderson & Lebiere, 1998b), subjects for the study were novices who had not yet been introduced to the alphanumeric symbol system of algebra. This study examined the character of their algebraic competence immediately following an initial exposure to algebra rules.

We should point out that this use of experimental methods for evaluation of theoretical hypotheses is remote from so-called *horse-race* research in which the relative efficacy of one curriculum is compared to that of another (Schoenfeld, 1987). The instructional methods and materials used in this study are not "curricula" in the educational sense. They are experimental treatments devised to isolate variables of interest while controlling for extraneous variables. We were not concerned with replicating normal or realistic instructional practices, nor are we putting forth these instructional methods or materials as educationally useful or interesting. Rather, we are making more basic claims about the nature of students' cognitive engagement with the alphanumeric symbol system of algebra. It is from these claims that we derive interpretations and proposals for algebra curriculum.

Subjects

One hundred fourteen Grade 7 students (generally about 12 years old) in four intact classes participated in the study. Two classes each were drawn from a predominantly middle class middle school (School 1) and a university laboratory school (School 2) in Baton Rouge, Louisiana. This grade level was chosen because algebra had not yet been introduced. Specifically, rules for transforming algebraic expressions had not previously been studied. Students whose last six week mathematics grade was A, B, or C were categorized as high achievers. Students receiving a grade of D or F were classified as low achievers.

Treatments

This study involved a two-treatment teaching experiment. Students in their intact classes were presented with 50 minute lessons for two consecutive days during their usual mathematics class period. A third classroom period included a 30 minute review lesson followed by a posttest, with an unannounced retention test following one week later. The lessons focused on eight rules for transforming algebraic expressions such as the *difference of squares* rule and the *distributive* rule. In one class at each school the notation used to represent expressions and rules was the standard (*ordinary*) alphanumeric notation of symbolic elementary algebra. For the other class a *tree notation* adapted from linguistic theory and artificial intelligence (Bundy & Welham, 1981; Kirshner, 1987; Thompson & Thompson, 1987) was used for all instruction, discussion, and testing. The distribution of students in the sample arranged by school, gender, mathematics achievement level, and the two treatments is displayed in Table 4.

Table 4

Distribution of Students in Sample by School, Gender, Mathematics Achievement, and Treatment

	School 1				School 2			
	Males		Females		Males		Females	
	High	Low	High	Low	High	Low	High	Low
Ordinary notation	13	5	7	3	10	8	7	3
Tree notation	7	7	8	8	7	7	11	3

Tree notation expresses the hierarchy of operations in an expression through the vertical arrangement of nodes. As well, in tree notation letters are used to represent operations that may be indicated only tacitly by positioning of symbols in ordinary notation. For instance, as shown in Figure 1 the expression $2x + 3$ would be represented as a syntactic tree structure in which addition (A), the least precedent operation, is at the top, with multiplication (M), below it. Tree notation was intended to provide a neutral medium to introduce algebraic symbol skills. Because visual salience is an artifact of the positioning and spacing of symbols in standard notation, we sought to use tree notation to view students' performance in learning algebra rules absent that factor.

A

M

2 x 3

Figure 1. Syntactic tree for $2x + 3$

Rules

Eight rules, which appear in Table 5, were used in the study. The four rules shown in the left column were selected because of their greater visual salience; the other four rules in the right column were selected because of their lesser visual salience. In order to help ensure the two rule sets did not vary in other respects, care was taken to balance a number of extraneous characteristics including number of constant terms, number of parentheses, and number of operations. To achieve this balance, specialized versions of ordinary algebra rules were used in some instances. For example, $2(x - y) = 2x - 2y$ was used in place of the more general $w(x - y) = wx - wy$ to equalize the number of numeric versus alphabetic characters across rule type. The rule sets were constructed such that both groups contained a total of eight constants. The total number of operations was nearly balanced: 20 for the visual rules; 21 for the non-visually-salient rules. The visual rules incorporated a total of five sets of parentheses on the left-hand sides of equations, whereas the non-visually-salient rules contained four sets of parentheses on the left-hand sides. An arbitrary order of presentation of the eight rules was used consistently across the two treatment groups.

Table 5

Rules Used in the Study

Visually salient Rules	Non-visually-salient rules
$2(x - y) = 2x - 2y$	$x^2 - y^2 = (x - y)(x + y)$
$(xy)^2 = x^2y^2$	$(x - y) + (w - z) = (x + w) - (y + z)$
$(x^y)^{-1} = x^{y \cdot -1}$	$(x - 1)^2 = (x^2 - 2x) + 1$
$(\frac{x}{y})(\frac{w}{z}) = \frac{xw}{yz}$	$x(y^{-1}) = \frac{x}{y}$

Instruments

The posttest consisted of 24 multiple-choice items. Each test item presented students with an algebraic expression. The task was to select from among six choices (including “none of these”) the expression that would result from application of one of the rules to the given expression. The available choices included expressions similar to the correct answer. For example, in ordinary notation, the expression $(7x)^2$ had response choices $(7^x)^2$, $7x^2$, 7^2x^2, 7^{x2}, 7^2x, and “none of these”. Figure 2 presents the same item in syntactic tree notation.

P

M = ?

7 x 2

1	2	3	4	5	6
P	M	M	P	M	none
P	P	P P	M	P	of
7 x 2	7 x 2	7 2 x 2	7 x 2	7 2 x	these

Figure 2. A test item and answer options in syntactic tree notation.
Note. M represents multiplication; P represents exponentiation or "power".

Three items were presented for each rule. Two of the three items were *recognition* tasks and the other item was a *rejection* task. In recognition tasks, like that shown in Figure 2, a correct answer could be found among the five multiple choice options. In rejection tasks no rule applied to the given expression, and the correct answer was "none of these." Among the other choices for rejection tasks was one in which a rule nearly applied. For example, in a rejection item (not shown) that used the expression $x^2 + y^2$, none of the rules students learned could be applied, although the choice $(x - y)(x + y)$ might be selected as an answer if the *difference of squares* rule were overgeneralized to a *difference-or-sum of squares* rule. Thus, recognition tasks assessed a student's ability to identify routine applications of the rules; rejection tasks evaluated a student's ability to constrain overgeneralizing the context of application of the algebra rule. The 24 test items were presented in random order.

A retention test given one week later was identical in structure to the posttest, except that variables and noncritical constants were changed (e.g., $(5w)^2$ = ? replaced the posttest question $(7x)^2$ = ?). The two treatment groups received identical tests, except for the notational form in which items were presented. Both posttest and retention test results were computed as the percentage correct. Only students who participated in the two instructional sessions were allowed to take the tests. Nine students present for the posttest were absent for the retention test.

Procedures

Great care was taken in the instructional process to minimize extraneous differences in the two treatments. Particularly for ordinary notation in which many of the operations are represented by the positioning of symbols (e.g., juxtaposition for multiplication, diagonal placement for exponentiation) rather than by an explicit symbol (e.g., M for multiplication, P for exponentiation, or power) it would be easy to slip into a more informal mode of exposition. For example, "3*x*" might be read as "three ex," or "x^3" as "ex cubed." To counteract this tendency, a specified lexicon was used for all instructional references to the operations. For instance, multiplying *a* by *b* always was stated as "*a* multiply *b*" or as the "multiplication of *b* by *a*" regardless of whether the notation used was tree notation or the standard alphanumeric system.

Parsing is another aspect of instruction that required careful consideration. In tree notation, the parse of the expression is built into the structure of the branching configuration. Thus in Figure 1 the precedence of multiplication over addition is indicated by the fact that only the M actually connects literals (2 and *x*) through its branches. There is no such forced interpretation in ordinary notation, so the equivalent expression $2x + 3$ could, in principle, be interpreted as $(2x) + 3$ or as $2(x + 3)$. It is merely notational convention that holds multiplication as more precedent than addition when parentheses are missing. This could have introduced the need for substantially different instructional treatments for the two groups, except that Kirshner (1989) noticed that students initially learn to parse algebraic expressions by "reading" the spacing and positioning cues as syntactic markers. Thus students learn to make the appropriate interpretation of $2x + 3$ not because of explicit instruction about operation levels, but because close spacing comes to signify greater precedence than wide spacing. We capitalized on this learning tendency by saying very little about parsing structure in either instructional treatment. In both cases, we just let the symbol system "speak for itself." We supplemented this strategy by a somewhat more liberal usage of parentheses in ordinary notation than is strictly necessary. For example, the rule $(x - 1)^2 = (x^2 - 2x) + 1$ used an extra set of parentheses on the right-hand side of the equation so that students would not have to master the left-to-right parsing convention for cases where operations are of the same *level*, as for example with addition and subtraction. As well, we minimized the syntactic complexity of all examples used in instruction and testing. For instance, the *difference of squares* rule was illustrated and tested through examples like $6^2 - w^2 = (6 - w)(6 + w)$ rather than through syntactically more complex examples like $(6v)^2 - w^2 = (6v - w)(6v + w)$.

All classes were taught by the second author. On the first day of instruction several expressions were presented to each class to introduce students to the notation with which they would be working. Then, each of the eight rules was taught individually using four examples, as follows. First, the rule was displayed on an overhead projector, and the teacher gave a declarative explanation of the rule. For instance the rule $x^2 - y^2 = (x - y)(x + y)$ was explained as "when the left hand side consists of a number, power 2, subtract another number, power 2, the right hand side will be the first number subtract the second number multiply the first number add the second number." Traditional names for rules (e.g., *difference of squares*) were not used. Next, an example involving other literals was used to illustrate the rule. With the rule and first example still visible, the left-hand side of a second example was shown and a student was arbitrarily chosen to predict the result. Other students were called upon to support or challenge the answer given. For the final two examples, the rule and the previous examples were removed from the overhead. The instructional procedure was the same as for the second example, except students no longer had the benefit of being able to see the rule.

Instruction was entirely procedural in that the technical process of rule application was stressed without mention of the numerical or referential meaning of the rules.

During the second day of instruction all eight rules were shown together, after which the rules were reviewed individually using the same format as on the previous day. Then, a 17 item multiple choice practice quiz was given featuring the sorts of items to be used in the tests. After completing the practice quiz, the class as a whole reviewed the items one-by-one, with discussion. The third day was identical to the second day, except the practice quiz was replaced with a 24 item posttest of similar form. The posttest was not reviewed in class. One week later, the students were given a retention test of the same form as the posttest. They had not been told to expect this test.

A nonparticipant observer was present for the three teaching days to verify that teaching protocols were adhered to and that preferential treatment was not given to either group or to either rule type, and to check for any differences in the dynamics and organization of the classes that might discriminate outcomes. The observer reported no such biases.

Analysis

The purpose of the study was to investigate the role of visual salience in the initial learning of algebra. If the initial learning of rules in algebra is acquisition of declarative information (as postulated in traditional information processing theories), then the results for visually salient rules should match those for non-visually-salient rules. For instance, students might learn the rule $2(x - y) = 2x - 2y$ as indicating that when a difference of two values is doubled the result is the same as when the individual terms are doubled and then the difference between them is found. For this kind of declarative representation the visual presentation of the rule should not be of consequence. This is because compilation of declarative knowledge into procedural knowledge generally is understood to occur over long periods of engagement within a domain, not in initial engagements such as occurred in this study. (For example, Anderson, 1983, studied compilation of geometry knowledge through thirty 45-minute lessons.) However, if algebra learning does not always begin with declarative knowledge, then visual salience may immediately begin to influence the character of students' performance.

Performance on recognition and rejection tasks explored this question in complementary ways. Recognition tasks test the superficial pattern recognition that might be expected to occur if the learner is attending to visual form rather than structural meaning. In this case, students should demonstrate more successful learning of visually salient rules than of non-visually-salient ones. Rejection tasks require the learner to resist overgeneralizing the domain of application of rules to apparently similar situations. In this case attending to form rather than to meaning would hurt performance. On the other side of the coin, if learning really begins with declarative understanding there should be no systematic differences between visually-salient rules and non-visually-salient rules on either type of task. Tree notation provides a kind of base line. In this notation, visual salience (which is an artifact of the positioning of symbols in standard notation) is eliminated. Thus, the tree diagram notation was intended as a means to compare the learning of the two sets of rules without the possible influence of visual salience.

Recognition and rejection tasks were analyzed separately by using analysis of variance for repeated measures. The analysis employed a $2 \times 2 \times 2 \times 2 \times 2$ factorial design including four between factors: school, gender, mathematics achievement, and notation (treatment); and one repeated within factor: rule type. A supplementary analysis of item responses was performed in addition to the subjects analysis to see if the results were general across the rules used in the study. This analysis used a similar factorial design, but included rule type as a between factor repeated over the within factors of school, gender, mathematics achievement, and notation treatment.

RESULTS

Recognition Tasks

The posttest analysis for within subjects effects indicated a significant rule type effect, $F(1, 97) = 17.07$, $p < .0001$, and a significant interaction between rule type and treatment, $F(1, 97) = 36.82$, $p < .0001$. The rule type by treatment interaction was also found to be significant in the items analysis, $F(1, 6) = 25.47$, $p < .005$. A table of mean percentages on the posttest and retention test for the rule type by treatment interaction is provided in Table 6.

Table 6

Recognition Task Posttest and Retention Test Percentages by Rule Types and Treatments

	Posttest		Retention test	
	Visually salient	Non-visually-salient	Visually salient	Non-visually-salient
Ordinary notation	73.0 (56)	40.4 (56)	68.5 (52)	35.4 (52)
Tree notation	48.9 (58)	55.8 (58)	47.4 (53)	53.1 (53)

Note. Values in the parentheses represent the number of subjects within the cell.

A test of simple effects indicated the treatment group using ordinary notation performed significantly better, $F(1, 111) = 28.53$, $p < .0001$, on the visually salient rules, but significantly worse, $F(1, 111) = 7.72$, $p < .01$, on the non-visually-salient rules than the tree notation treatment group. In accordance with our hypothesis, recognizing visually salient rules was significantly easier (73.0% correct) than recognizing non-visually-salient rules (40.4% correct) for students taught and tested in ordinary notation, $F(1, 56) = 4.14$, $p < .05$. For students using tree notation, scores on visually salient and non-visually-salient rules seem about equal (48.9% and 55.8%, respectively). Interestingly, however, the slight *reverse* tendency for visually salient rules to be harder (48.9% correct) than non-visually-salient rules (55.8% correct) also was significant, $F(1, 55) = 77.72$, $p < .0001$. As we had balanced the two rule sets as nearly as possible on characteristics other than visual salience, we can offer no explanation for the differences in difficulty discovered for tree notation. However, if anything, these differences highlight the greater base line difficulty of the visually salient rule set, making the converse findings in ordinary notation more pronounced.

The results of the subjects analysis on the retention test were similar. The within subject effects indicated a rule type effect, $F(1, 89) = 21.12$, $p < .0001$, and a rule type by treatment interaction, $F(1, 89) = 44.97$, $p < .0001$. The interaction was also found in the items analysis, $F(1, 6) = 36.43$, $p < .001$.

A test of simple effects for the retention test indicated students using ordinary notation performed significantly better, $F(1, 103) = 17.80$, $p < .0001$, on the visually salient rules, but significantly worse, $F(1, 103) = 11.22$, $p < .0001$ on the non-visually-salient rules than the students using tree notation. In contrast to the posttest, no significant difference between the two rule types was found for students using tree notation (47.4% and 53.1%, respectively). But for students using ordinary notation, visually salient rules were significantly easier (68.5%) than non-visually-salient rules (35.4%), $F(1, 51) = 91.66$, $p < .0001$.

Other significant effects were found on both tests that do not bear directly on the central hypothesis. A school effect was found on the posttest with School 2 subjects having a higher mean percent correct (63.6%) than School 1 subjects (45.5%) in the subjects analysis, $F(1, 97) = 19.72$, $p < .0001$, as well as in the items analysis, $F(1, 6) = 54.27$, $p < .001$. An effect of past mathematics achievement was also found on the posttest, $F(1, 97) = 12.05$, $p < .001$, with high achieving students scoring better (61.1%) than their low achievement counterparts (43.8%) in the subjects analysis, as well as in the items analysis, $F(1, 6) = 12.32$, $p < .05$.

The retention test indicated similar trends. A school effect was found with School 2 subjects scoring significantly better (58.8%) than School 1 subjects (43.5%) in the subjects analysis, $F(1, 89) = 13.89$, $p < .0005$, as well as in the items analysis, $F(1, 6) = 6.74$, $p < .05$. In addition high achievement students scored better (56.4%) than low achievement students (41.6%) in the subjects analysis, $F(1, 89) = 8.84$, $p < .005$, as well as in the items analysis, $F(1, 6) = 17.43$, $p < .01$.

Gender effects. The posttest analysis for within subjects effects also indicated an interaction between rule type and gender, $F(1, 97) = 4.30$, $p < .05$. These results are shown in Table 7. A test of simple effects indicated male subjects performed significantly better (55.4%) than female subjects (42.6%) on the non-visually-salient rules, $F(1, 111) = 5.18$, $p < .05$. Whereas no significant difference was found between rule types for male subjects, female subjects performed significantly better on the visually salient rules (62.8%) than on the non-visually-salient rules (42.6%), $F(1, 63) = 23.99$, $p < .0001$.

Table 7

Posttest Recognition Task Percentages for Rule Types By Gender

	Visually Salient	Non-visually-salient
Male	58.4 (64)	55.4 (64)
Female	62.8 (50)	42.6 (50)

Note. Values in the parentheses represent the number of subjects within the cell.

In an attempt to interpret these data, we examined the more specific means for male and female students in the two treatment groups. The results from this analysis appear in Table 8.

Table 8

Posttest Recognition Task Percentages for Rule Types By Gender and Treatment

	Ordinary notation		Tree notation	
	Visually salient	Non-visually-salient	Visually salient	Non-visually-salient
Male	70.6	48.8	50.0	59.9
	(36)	(36)	(28)	(28)
Female	74.3	35.8	47.8	53.1
	(20)	(20)	(30)	(20)

Note. Values in the parentheses represent the number of subjects within the cell.

Although the rule by gender by treatment interaction was not significant, $F(1, 97) = 2.03$, $p = .16$, we can see the trends in the data that contribute to the overall significant effects when collapsed across treatment groups. In tree notation male students tended to excel on the non-visually-salient rules (59.9% correct), whereas in ordinary notation female students tended to lag on the non-visually-salient rules (35.8% correct). The combination of these tendencies appears to constitute the significant overall effect we found. Because we designed tree notation to eliminate visual salience, and because we designed the visually salient and non-visually-salient rule sets to be equivalent in all other respects, we can offer no explanation of the first of these tendencies in the data. However, the second tendency mirrors results in a previous study where it was suggested that a field independent cognitive style might make male students less dependent on the visual arrangement of symbols in ordinary notation than female students (Kirshner, 1989). Because the interaction effect was not significant, we will not pursue such speculations further in this article.

Rejection Tasks

In ordinary algebra notation the visual salience of certain rules that makes them easier to recognize may mediate against developing declarative representations. If so, it was anticipated that subjects would be less able to constrain overgeneralizing these rules in comparison with rules that lack visual salience. For syntactic tree notation, since subjects must engage equally with the declarative content for both rule types, no significant differences were expected between rule types.

A rule by treatment interaction was indicated by the within subjects effects in the subjects analysis, $F(1, 97) = 11.31$, $p < .005$, and marginally in the items analysis, $F(1, 6) = 5.12$, $p < .065$. Posttest and retention test mean percentages for the rule type by treatment interaction are provided in Table 9.

A test of simple effects for the posttest indicated the ordinary notation treatment group performed significantly better, $F(1, 111) = 14.57$, $p < .0005$ than the tree notation treatment group on the non-visually-salient rules. In tree notation it was easier to constrain overgeneralization for visually salient rules (15.2% correct) than for non-visually-salient rules (10.9% correct), $F(1, 56) = 5.92$, $p < .05$, but in ordinary notation non-visually-salient rules were easier to constrain (19.8%) than visually salient rules (13.0%), $F(1,55) = 11.94$, $p < .005$. As with the recognition tasks, we have no explanation for the differences in difficulty of rule types given in the tree notation. But as with recognition tasks these differences in base line difficulty contrast with, and hence emphasize, the differences emerging in ordinary notation.

Table 9

Rejection Task Posttest and Retention Test Percentages By Rule Types and Treatments

	Posttest		Retention test	
	Visually salient	Non-visually-salient	Visually salient	Non-visually-salient
Ordinary notation	13.0 (56)	19.8 (56)	8.3 (52)	15.9 (52)
Tree notation	15.2 (58)	10.9 (58)	13.0 (53)	7.4 (53)

Note. Values in the parentheses represent the number of subjects within the cell.

As can be seen in Table 9, the pattern of results in the retention test matches that of the posttest. Whereas in tree notation subjects found the visually salient rules to be easier to constrain (13.0%) than the non-visually-salient rules (7.4%), in ordinary notation the visually salient rules were more difficult to constrain (8.3%) than the non-visually-salient rules (15.9%). However, basement effects may have reduced our ability to detect significant effects here. The items analysis was not significant. The within-subjects effects did indicate a rule type by treatment interaction, $F(1, 89) = 27.49$, $p < .0001$; however, a four-way interaction, $F(1, 89) = 7.19$, $p < .01$, involving school, gender, treatment, and rule type render these data uninterpretable. Finally, as with the posttest, an achievement effect was found on the retention test, $F(1, 97) = 6.39$, $p < .05$, with high mathematics achievement students performing better (34.1%) than their low achievement counterparts (22.1%) on the subjects analysis.

CONCLUSIONS

In this study, Grade 7 students not previously exposed to algebraic symbol manipulation tasks were taught to apply eight algebra rules over the course of two lessons. Half of the rules were selected for their high *visual salience*–a visual coherence that seems to make the left- and right-hand sides appear naturally related to one another (see Table 5, left column). The other four rules were selected for their low visual salience (see Table 5, right column). We compare the effect of visual salience to an animation sequence in which distinct visual frames are perceived as ongoing instances of a single scene. We see this effect as relieving the observer of any obligation they might otherwise feel to make conscious connections between two separate entities, as they might be inclined to do for non-visually-salient rules.

Two kinds of tasks were used to assess the character of students' initial knowledge of the algebra rules. Recognition tasks assessed the students' ability to identify routine applications of the rules. Rejection tasks provided an expression that could not be manipulated by any of the rules, thereby measuring a student's ability to constrain overgeneralizing the context of application of the given rules. Thus if students were engaging with rules based on the visual character of expressions we would expect visually salient rules to be easier to recognize (recognition tasks) but harder to constrain (rejection tasks) than non-visually-salient rules.

The subjects and items analyses for the posttest support the hypothesis that students engage with the visual characteristics of the symbol system in their initial learning of algebra rules. Percentage correct scores for recognition tasks were significantly higher for visually salient rules than for non-visually-salient rules. Such scores for rejection tasks were significantly lower for the visually salient rules. The tendency for the results on the retention test were similar; however, with the overall degradation of performance over time, only some of these differences retained statistical significance.

To the extent possible we tried to balance the two groups of rules with respect to other variables that might account for differential results (number of constant terms, number of parentheses, number of operations). But we went further than this. In order to obtain baseline data on rule difficulty independent of visual salience, half of the subjects were taught using a tree notation that displays the hierarchy of operations in an expression through the vertical arrangement of nodes. Tree notation was used because it presents an alternative medium that does not exhibit the visual salience present in the usual notation. For example, consider the visually salient rule $(\frac{x}{y})(\frac{w}{z}) = \frac{xw}{yz}$ and the non-visually-salient rule $(x - y) + (w - z) = (x + w) - (y + z)$

both used in the study. What may be masked by the visual characteristics of ordinary notation is that the two rules are structurally identical, except for the *level* of the operations: Multiplications and divisions in the first rule are systematically replaced by additions and subtractions in the second rule. In tree notation, where visual salience does not dominate perception, the structural similarity is plainly evident, as shown in Figure 3.

```
        M                    D                    A           S

     D     D   =        M     M                     S     S   =      A     A

    x y   w z          x w   y z                   x y   w z        x w   y z
```

Figure 3. Comparing tree notation for $(\frac{x}{y})(\frac{w}{z}) = \frac{xw}{yz}$ and $(x - y) + (w - z) = (x + w) - (y + z)$

Because of our efforts to balance other variables that might account for differences, we anticipated that tree notation would demonstrate the underlying equal difficulty level of the two rule sets. Contrary to our expectation visually salient rules turned out to be significantly *more* difficult to recognize and significantly *less* difficult to constrain than non-visually-salient rules when presented using tree notation. We have not found any convincing explanation for these inversions of results; but if anything they emphasize that the baseline difficulty of the two rule-sets mitigated *against* detecting the influence of visual salience, so conspicuously present in this study.

Given Davis' (1984) previous identification of what he called *visually moderated sequences* in which a visual cue elicits a procedure which produces a new visual cue which elicits a new procedure, and so on, the idea that processing of algebraic symbols involves visual cues cannot be considered novel or surprising. However, the assumption of previous cognitive science research that declarative knowledge comes first (Anderson & Lebiere, 1998b) is challenged in this study. By selecting younger students who had not previously encountered algebra transformation rules, we have demonstrated that visual pattern matching is immediate and spontaneous. Thus, this study reopens questions about students' experience in traditional algebra instruction that have long been considered settled, and renews possibilities for a structural algebra curriculum that have largely been abandoned by mathematics education theorists.

DISCUSSION

The directions of current reform efforts in algebra education are firmly set within the NCTM *Principles and Standards for School Mathematics* (2000):

> In general, if students engage extensively in symbolic manipulation before they develop a solid conceptual foundation for their work, they will be unable to do more than mechanical manipulation (National Research Council [NRC], 1998). The foundation for meaningful work with symbolic notation should be laid over a long time (p. 39).

The source of this "solid conceptual foundation" was summarized by the NCTM's Algebra Working Group (1998) which charted a Framework for reform:

> The "Framework" proposes a way to develop algebraic reasoning by exploring a variety of contextual settings that are connected by organizing themes. By serving as organizers, themes help students recognize important ideas and make connections. Contextual settings are the ground on which these themes play out. They provide the substance from which and about which to reason. (p. 164)

NCTM's conclusion that work with algebraic symbols can only become productive for students within contextually rich settings summarizes a consensus in the algebra education research community that has been building for several decades. The foundation upon which that consensus rests is a shared understanding of the failure of the traditional curriculum as manifest in the seemingly mindless errors patterns of the sort listed in Table 1. For instance Booth's (1989) conclusion that "algebraic representation and symbol manipulation ... should *proceed from* an understanding of the semantics or referential meanings that underlie it" (p. 58), followed upon her extensive study of students errors (Booth, 1984). Similarly, Fey (1992) led development of the contextually rich Computer-Intensive Algebra curriculum, in part, because "many students do not become proficient in the skills of algebra ... [and] very few students acquire the understanding of algebraic ideas and methods that is required to reason effectively with symbolic expressions" (p. 1). And Kaput (1995) based his call for contextually rich curricula on "the current wholesale failure of school algebra:"

> acts of generalization and gradual formalization of the constructed generality must precede work with formalisms – otherwise the formalisms have no source in student experience. The current wholesale failure of school algebra has shown the inadequacy of attempts to tie the formalisms to students' experience *after* they have been introduced. It seems that, "once meaningless, always meaningless." (pp. 74-75)

The conclusion of the research community that algebraic concepts need to be explored in contextually rich settings follows reasonably from the interpretation of persistent errors as indicating students' failure to manage the declarative demands of abstract, decontextualized rules. But the data reported here challenge this interpretation. Students appear to begin their processing of (at least) some rules, presented in a didactic manner, at the *subcognitive* level (Hofstadter, 1985) of visual pattern matching rather than just at the conscious level of declarative content. Thus students' persistent algebra errors may stem from disengagement from declarative content rather than from difficulties with the declarative content, per se.

The study reported here, by itself, does not provide evidence of the extent to which visual salience may be an influence upon students' algebra learning. However, in the remainder of this section we integrate the results of this study with those of a previous investigation into the visual syntax of algebraic expressions (Kirshner, 1989) to produce a more comprehensive account of algebra skill development through traditional instruction. This account reveals the visual pattern modality as broadly dominating algebra skill development. Thus we believe previous researchers have been premature in concluding from the persistence of seemingly mindless procedural errors that students cannot handle the declarative content of elementary algebra rules. Rather, we believe that students capabilities in this regard have not been tested. In the next section we outline a curricular approach that enhances the declarative character of algebra by attenuating the effects of visually salient notations through introduction of a specialized lexicon. Such a curriculum has the possibility of re-energizing the traditional mathematics education interest in introducing secondary school mathematics students to algebra as a formal, structural endeavor, to complement the exciting possibilities for referential algebra that currently are being developed.

Retheorizing Error Patterns in Algebra Symbol Skills

The results here extend a previous study that examined the influence of visual cues in the syntactic parsing structure of algebraic expressions (Kirshner, 1989). That study was concerned with the nature of students' competence in identifying the syntactic structure of an expression like $5x^2$ which permits two possible parses: $(5x)^2$ and $5(x^2)$. The conventions for operation precedence (in the absence of parentheses or other signs of aggregation) can be neatly summarized in terms of *operation levels*

(a) Higher level operations are precedent

(b) If adjacent operations are of equal level, the operation on the left is precedent

where operation level is given a declarative interpretation:

Level 1	addition and subtraction
Level 2	multiplication and division
Level 3	exponentiation and radical

(Level 3 operations are said to be higher than Level 2 operations which are higher than Level 1 operations.)

Kirshner (1989) noticed that operation level also could be given a visual interpretation based on the spacing and positioning of symbols:

Level 1	wide spacing	$a \pm b$
Level 2	horizontal/vertical juxtaposition	ab or $\frac{a}{b}$
Level 3	diagonal juxtaposition	a^b or $\sqrt[a]{b}$

In this interpretation, deciding, for example, that $5x^2$ groups the x with the 2 prior to the 5 is accomplished through a visual hierarchy of diagonal juxtaposition ahead of horizontal juxtaposition rather than through a declarative hierarchy of exponentiation before multiplication. Kirshner (1989) demonstrated that for many students competence in parsing depends on the usual alphanumeric display. Declarative knowledge does not become sufficiently well established to enable correct parsing without the support of the visual relations in standard notation. Thus visual salience comes into play not only in the character of some transformational rules, but also in the parsing structure of algebraic expressions.

How are these two aspects of visual salience interrelated? A declarative account of algebra rules would depend on explicit knowledge of both the parsing structure of expressions and the transformations accomplished by each rule. For example, an explicit understanding of the rule $(xy)^2 = x^2y^2$ requires knowing the context of application, a product is raised to a power, as well as the action of the transformation, the power is distributed over the product. Visual salience of syntactic parsing structure enables students to work with the form $(xy)^2$ without engaging explicitly with its structural elements. Visual salience of transformations enables the action to be carried out without reflective grounding.

What is intriguing is the collusion of visual salience of transformational patterns and parsing cues in the production of error patterns like those displayed in Table 1. The errors students tend to make in overgeneralizing rules are related to syntactic structure: they overgeneralize the context of application of the rule, not the nature of the transformational action. For example when students overgeneralize $(xy)^2 = x^2y^2$ as $(x + y)^2 = x^2 + y^2$, they are overgeneralizing the context of application of the rule; the transformational action is essentially correct. As Davis and McKnight (1979) noticed in their extensive study of algebra errors: "The *syntax* of algebraic expressions may be a key or milestone kind of knowledge in algebraic learning. The degree of syntactic security seems to be a crucial element in a student's predisposition to regression under strain" (p. 56).

However, if students' errors are related to insecurities with the syntactic structure of expressions, why do these errors tend only to occur with visually salient rules? Why do students fail to overgeneralize, say, $x^2 - y^2 = (x - y)(x + y)$ as $x^2 + y^2 = (x + y)(x - y)$? In part, we suspect this may have to do with an anomaly in the algebra lexicon: The names for non-visually-salient rules tend to provide more support for explicitly apprehending the parsing structure of the input expression than do names for visually salient rules. For example, $x^2 - y^2 = (x - y)(x + y)$ is called "the difference of squares" rule, a title that identifies the syntactic form of the input expression, whereas $(xy)^2 = x^2y^2$ is not normally called "the power of a product" rule, which would likewise identify the syntactic form of the input expression. (Compare also "the sum or difference of cubes" rules, and the "binomial square" rule with unnamed visually salient rules like $(x^y)^z = x^{yz}$ and $x^{y+z} = x^y x^z$).

But our data lead us to believe the lexical issue is only a minor factor: we achieved these same overgeneralizations for visually salient rules while controlling for extraneous variables like differential naming of rules. Rather, we think visual salience of transformations influences performance by creating the illusion of an animation sequence: The perception that the left and right hand sides of the equation are ongoing instances of a single scene obviates the need to engage explicitly with the structural description of either expression. In contrast, the sense of dealing with separate entities in the case of non-visually-salient rules forces students to attend more fully to the structural descriptions. This explains why the rules that tend to become overgeneralized by students are visually salient.[4]

---

[4]This psychological argument does not obviate the need for further historical and ontogenetic analysis. As we reviewed in the introduction, mathematicians of Hamilton's day were inclined to accept Peacock's (1833) Principle of the Permanence of Equivalent Forms in which the generality of certain rules was accepted without argument or justification. It seems that the new notations being incorporated into algebra had the consequent effect of transforming *rules*—statements of specific relationships—into *patterns* that could be carried easily over to new domains. We offer, with speculative abandon, the proposal that as algebra progressed from its rhetorical and syncopated phases into its symbolic phase (Sfard, 1995), it became a repository in spatial form for grammatical patterns that normally play out in the temporal space of speech and language. For instance, the syntactic precedence of adjacent expressions over spaced expressions that resolves the ambiguity of $a + bc$ as $a + (bc)$ [rather than $(a + b)c$] has a counterpart in the parsing of natural language, as revealed in the resolution of the ambiguity of the oral expression /light/ /house/ /keeping/ through temporal juxtaposition, as either /lighthouse/ /keeping/ or /light/ /housekeeping/. Similarly, in the transformational realm, the distributive pattern present in some algebraic contexts (e.g., $c(a + b) = ca + cb$, [illegible]) but not others (e.g., $\sqrt[c]{a+b} \neq \sqrt[c]{a} + \sqrt[c]{b}$), is preceded, historically and ontogenetically, by vast experience in grammatical interpretation in natural language. For instance, "Old men and women were released first", is ambiguously interpretable either distributively as "Old men and old women were released first," or non-distributively as "Old men and [all] women were released first." Similarly, "I like cake and ice cream" is interpretable distributively as "I like cake and I like ice cream," or non-distributively as "I like cake and ice cream [together, but not separately]."

Unfortunately, rather than intervening to establish a declarative focus, the traditional algebra curriculum actively *supports* a reliance on visual processing. From one of our own secondary school algebra learning experiences we recall being taught the BOMDAS rule for order of operations: Brackets, Of, Multiplication, Division, Addition, Subtraction. This account of order of operations mixes a marker of aggregation, brackets (there are others as well that are not mentioned–the vinculum as used in $\dfrac{a}{b+c}$ or $\sqrt{a+b}$, and superscription as in a^{bc}), together with a hierarchy of operations that applies in the absence of aggregation markers. Moreover, the hierarchy of operations does not mention left-to-right precedence for operations of equal level and even omits exponentiation and radical, rendering the mnemonic unable to handle such straightforward cases as $5x^2$.

We believe such declaratively insufficient instruction endures in the traditional algebra curriculum because of a persistent misinterpretation on the part of educators of students' *successful* performances. That students are able to make correct parsing decisions with algebraic expressions like $5x^2$ is taken as evidence they have mastered the declarative content of the hierarchy of operation. Similarly, a quick competence with routine applications of transformational rules is taken to indicate a degree of declarative mastery of the rules. The visual patterning underlying these competencies is not appreciated. Thus instruction moves on to more complex skills before the declarative fundamentals have been established.

This premature abandonment of fundamentals tends to create momentum for the continued dissociation of skills from a structural understanding of algebra. Lacking the basic structural perspectives they need to reason explicitly about new rules and procedures they are learning students become increasingly reliant on visual pattern matching competencies. Eventually, many students do manage to persevere to a mute competence in algebra as visual pattern matching processes become

Thus the propensity to proliferate algebraic forms, historically and ontogenetically, may be conditioned by a broader syntactic capacity.

sufficiently refined to successfully constrain overgeneralizations. But this is a shallow victory, most often achieved without insight or interest, without connection to the broader projects of mathematics, and without providing a foundation for further theoretical studies or for insightful application. It is only a small minority of students, perhaps because of a field independent cognitive style (Kirshner, 1989), that seem to orient themselves to an explicit structural interpretation of algebra. Thus Kaput's (1995) above quoted dictum, "'once meaningless, always meaningless'" (p. 75) is correct. But meaninglessness in manipulating expressions and equations in traditional algebra instruction stems from an absence of declarative fundamentals, not from an absence of referential context.

A NEW PEDAGOGICAL DIRECTION

This analysis, we believe, points the way to a new pedagogical approach for elementary algebra, an approach that requires syntactic and transformational processes to be articulated declaratively, enabling more, rather than fewer, students to escape the notational seductions of nonreflective visual pattern matching. But first, we want to clarify that the pedagogical issue we engage within this article is not whether algebra instruction should make use of rich contextual settings. We take it as obvious and uncontested that algebra has vitally important application to both number and quantity, and that contextually rich settings are ideal for exploring such applications. Indeed, the possibility of escape from mindless symbol manipulation through referential context has been enabled by computational technologies that not only perform symbolic manipulations instantly but also can hot link equations, graphs, and tables with real world settings to provide extraordinary educational opportunities for students to experience algebraic relations dynamically (e.g., Romberg, Fennema, & Carpenter, 1993). Rather, we take issue with the belief that for many students such settings can provide the *only* route to algebraic meaning. Our overall curricular view is that the referential and abstract-decontextualized facets of algebra need to grow in tandem with one another, each approach having its own integrity and space in the curriculum, neither subsuming the other. We introduce a specialized lexicon as a new curricular approach to algebra as a formal, structural study to complement the current focus on referentially rich domains.

The approach we propose is to intercede discursively between the students' spontaneous matching of visual patterns and their performance of mathematical skills by instituting a Lexical Support System (LSS) through which students give more precise declarative accounts of algebraic structures and processes (Kirshner, 1998). The foundation of the strategy is rigorous structural description of algebraic expressions. The starting point is an explicit declarative account of the conventions for parsing algebraic expressions (order of operations), as given above. From here, the *principal operation* of an expression is defined as the least precedent operation according to the parsing rules (e.g., the principal operation of $5x^2$ is multiplication because exponentiation has higher precedence than multiplication). The *principal subexpressions* of the expression are the parts of the expression joined by the principal operation (e.g., 5 and x^2 are the principal subexpressions of $5x^2$). Recursively, each subexpression can itself be parsed yielding a complete structural description of an expression. Lexical items like *term* and *factor* that usually are used casually in classroom conversation can now be rigorously defined: *Terms* (*factors*) are the principal subexpressions of an expression whose dominant operation is addition (multiplication).[5]

This explicit structural description of expressions is preliminary to rigorous description of transformational rules in terms of their structural effects on expressions, as well as to rigorous characterization of standard tasks. For instance, *to factor* is to transform an algebraic expression

[5]An alternative instructional approach might be to impose a pedagogical notation like tree notation as a substitute for ordinary notation as a way to forestall visual pattern matching. We recommend against this practice for two reasons. First, as with ordinary notation, tree notation may provide for it's own sorts of salience thus far unexplored and unexamined (for instance as may be implicated in the own inscrutable tree notation results in this study). Second, ordinary notation seems to have been designed and/or developed to create certain visual economies of processing. We see it as more productive to guide students carefully in the use of this powerful tool rather than to postpone its usage and then face the added burden of translating back to ordinary notation.

whose principle operation is addition or subtraction into one whose principle operation is multiplication. In the service of such a goal, one explicitly examines the right and left hand sides of the available rules to determine which ones accomplish the appropriate effect, and of those, which can apply to the given circumstance. Thus strategies for deploying transformational rules for particular purposes flow out of an explicit focus on the syntactic structure of expressions.

This focus on structural descriptions carries with it a certain discursive discipline that is widely ignored in the traditional algebra curriculum where structural and transformational data are routinely conflated. Thus we speak of the factors of $x^2 - y^2$ when what we literally mean is factors of another expression, $(x-y)(x+y)$ that is transformationally equivalent to this one. The expression $x^2 - y^2$ has no factors as its principal operation is subtraction rather than multiplication. The LSS curriculum requires this kind of literalism to maintain a declarative focus against the seductive tendency to match visual patterns.

The following contrived episode illustrates the sort of communicational possibilities opened up by these more rigorous discursive practices. This interaction, similar to many the first author has engaged in when using these methods instructionally, involves a student's erroneous cancellation of the 3s in $\frac{3x^2 + 1}{3y - 2} = \frac{x^2 + 1}{y - 2}$.

Teacher: What rule are you using in this step?

Student: The cancellation rule for fractions.

Teacher: Can you remind me what that rule is?

Student: It's the rule that allows canceling a common factor of the numerator and denominator of a fractional expression.

Teacher: Okay, let's take a look at it. What have you canceled?

Student: The threes, because they're factors, they're multiplied.

Teacher: Good, they are indeed factors, but are they factors of the *numerator and denominator*? Let's check. What is the principal operation of the numerator?

Student: Let's see, there's an exponentiation, a multiplication, and an addition. So the principal operation is addition, the least precedent one according to the hierarchy of operations.

Teacher: Good, now what are the principal subexpressions called in this case?

Student: They're called terms. ...Oh, I see, it has to be a factor of the *whole* numerator and denominator to be canceled; not just part of it.

Such communicative possibilities can be contrasted with traditional algebra instruction in which students and teachers talk past each other as they use words like "term" and "factor" without structural grounding. Perhaps the teacher admonishes the student to make sure they are canceling *factors*. But the structural distinction, so clear and tangible for the teacher, is not conveyed to the student. Instead, the student learns only that they have done something wrong and need to do something different. Absent an understanding of the structural fundamentals, what gets recorded for the student is something about the visual shape of incorrect and correct applications. Eventually, with persistence, the visual pattern matching processes becomes sufficiently refined as to constrain incorrect applications. In this way, what begins as an opportunity for communication of structural information is reduced to support for mindless matching of visual patterns.

Or perhaps the teacher, in the spirit of reform, asks the student to substitute values for variables to see the falsity of their transformations in numerical domains; or finds a geometric area model to illustrate geometric interpretations of algebra. In themselves, these activities are wholly laudable. But as a substitute for dealing with the structural complexities of algebra, the segue into referential context needs to be recognized for what it is: abandonment of the agenda of structural algebra. For it is not obvious how an agenda of contextually rich applications ever could produce more than fragments of deductive rigor. A true agenda of structural algebra requires mastering a sustained discourse of decontextualized, abstraction. As Bell (1936) put the structural case:

> The very point of elementary algebra is simply that it *is* abstract, that is, devoid of any meaning beyond the formal consequences of the postulates laid down for the marks. ... We miss the whole point of algebra if we *insist* on any particular interpretation. Algebra stands upon its own feet as a "hypothetico-deductive system." (p. 144)

THE LEXICAL SUPPORT SYSTEM IN HISTORICAL AND PHILOSOPHICAL PERSPECTIVE

Bell's (1936) pronouncements anticipated the "new math" movement of the 1960s and 1970s–the last concerted effort to orient the algebra curriculum around formal mathematical structure. However, critics of the new math agenda complained that its programs were "excessively formal, deductively structured, and theoretical.... [and] fail to meet the needs for basic mathematical literacy of average and low ability students" (National Advisory Committee on Mathematics Education [NACOME], 1975, p. ix). In this final section, we contextualize our own approach by distinguishing it from the new math curriculum, and by noting limitations of elementary algebra as a structural enterprise when viewed from a non-representational epistemological perspective.

The New Math

Despite the shared general focus on structural algebra, our Lexical Support System has a substantially different orientation from the new math. The new math curriculum was oriented by "the concepts of set, relation, and function and by judicious use of broadly applicable mathematical processes like deductive reasoning and the search for patterns" (Fey & Graeber, 2003, p. 524). As they explain:

> Much of the energy in proposals to accelerate and deepen student learning by emphasizing unifying concepts and structures of mathematics was drawn from developments in advanced mathematics, where a similar structural point of view was leading to unification and generalization of traditional branches of the subject through focus on pervasive abstract patterns. (p. 525)

To clarify the differences between the new math approach and our Lexical Support System, we need to distinguish between two foundational aspects of modern mathematics: logicality and formality/abstraction (Ernest, 1998). Logicality relies on explicit processes of inferential reasoning; formality, on rigorous application of uninterpreted rules.

The new math, as a departure from traditional algebra instruction, was notable for its logicist intentions (Ernest, 1985). Indeed, an explicit intention of the new math was to distribute part of the emphasis on deductive reasoning from the geometry curriculum to algebra:

> One way to foster an emphasis upon understanding and meaning in the teaching of algebra is through the introduction of instruction in deductive reasoning. The Commission [on Mathematics] is firmly of the opinion that deductive reasoning should be taught in all courses in school mathematics and not in geometry alone. (College Entrance Examination Board, 1959, p. 23)

This logicist orientation of the new math curriculum may have been a major cause of its difficulty for the general secondary student. Inferential reasoning is notoriously difficult for adults, let alone adolescents (Evans, 1982). Logical inferences begin with a conditional statement consisting of an antecedent and a consequent, one of which is asserted or denied in a subsequent statement. For instance the logical principle modus ponens asserts the conditional, *if p then q*, and the antecedent, *p*, from which one may deduce the truth of the consequent, *q*. There are three other inferential possibilities starting with the conditional, *if p then q*: denial of the antecedent (assert *not p*, deduce *not q*); affirmation of the consequent (assert *q*, deduce *p*); and *modus tolens* (assert *not q*, deduce *not*

p). However, only *modus ponens* and its contrapositive *modus tolens* are logically sound. The other two are common inferential errors asserted by a plurality of adults in various studies of conditional reasoning (Evans, 1982).

In contrast to the new math, the LSS focus is on the formal rigor of algebraic derivation–the processes whereby structural analyses are performed on expressions and equations, and transformational rules applied according to explicit syntactic conditions. Now, there is a sense in which an algebraic derivation can be accounted as a proof of equivalence. For instance, the derivation $3x^2 - 27 = 3x^2 - 3{\cdot}9 = 3(x^2 - 9) = 3(x^2 - 3^2) = 3(x - 3)(x + 3)$ demonstrates the equivalence of $3x^2 - 27$ with $3(x - 3)(x + 3)$. But in its logical structure, such derivations rely on biconditional reasoning rather than conditional reasoning: $3x^2 - 27$ is true if and only if $3(x - 3)(x + 3)$ is true; and each step of the derivation is logically reversible. For biconditional reasoning, all four of the inferential possibilities discussed in the preceding paragraph are valid. Thus the requirement for logical sophistication in such structural derivations is averted. It is only in relatively infrequent special cases (e.g., multiplying by 0), that arguments are not bi-directional, and sophisticated conditional reasoning demands arise. In this respect, the Lexical Support System approach is less ambitious than was the new math. However, it may also prove a more tractable curriculum for the general secondary student.

Limitations of Elementary Algebra as a Structural Domain

There is another important sense in which the LSS approach departs from the new math curriculum of the 1960s and 1970s. The new math intended to introduce algebra not only as a foray into the logical methods of mathematics, but also to reflect the explicit content of abstract algebra as a structural study of the number systems (e.g., Haag, 1961). Axioms of the number systems were explicitly introduced as foundational to the subsequent transformations and manipulations (Osborne & Kasten, 1992). However, this involved something of a sleight of hand with respect to exponentiation rules. Reasonably, the new math never attempted to explore with secondary school students the notions of limit or least upper bound that come from analysis rather than algebra and that are necessary for a rigorous characterization of the real numbers as a complete ordered field (Royden, 1968). Instead, the formal treatment was restricted to the field properties (and existence of roots of certain polynomials). A rigorous treatment of exponentiation is impossible from a purely algebraic point of view.[6]

Our LSS approach cuts the Gordian knot that bound the new math curriculum to theoretical *content,* by addressing structural algebra purely as mathematical *method*. Thus we are more eclectic in the rules we are prepared to introduce, not restricting ourselves to field axioms, but including exponent rules as well. Rather than addressing both content and methods of mathematics, our curricular focus seeks only to *enculturate* students (Kirshner, 2002) to mathematical "habits of mind" (Cuoco, Goldenberg, & Mark, 1996) by establishing a formal discursive classroom practice involving explicit structural analysis of algebraic expressions and equations.

On an ontological level, this departure from mathematical content accommodates the fact that secondary school algebra developed historically as a system for manipulating expressions and equations prior to the formalization of abstract algebra in the latter half of the 19th century (Kline, 1980). On an epistemological level, our non-representational view of cognition prepares us for the fact that cognitively, the discursive practice of formality may be more an idealization than an actual map of the processes involved in manipulating algebraic symbols. Although many aspects of elementary algebra symbol manipulation are easily captured by formal rule structures, Kirshner (2001) argued attempting to force it wholly into such a mold results in unacceptably contrived

---

[6]Indeed, the exponential function cannot even be defined in the reals without using some limit process: (What is 2^x ?). Interestingly, there is a branch of mathematics known as *real algebra* that seeks to find a set of algebraic (first order) axioms for exponentiation in the reals. But as Macintyre (1979/1980) observed, this project is far from concluded:

> The most interesting problem provoked by the above is that of showing that there are no "exotic" laws, i.e. that every law is a consequence of the laws of +, x, -, $^{-1}$, 0, 1 together with $x^1 = x$, $x^{y+z} = x^y x^z$, $x^{yz} = (x^y)^z$, $(xy)^z = x^z y^z$. It seems difficult to prove such a theorem by the methods of real algebra used above. (p. 197)

variations that are psychologically and pedagogically implausible. For instance a consistent binary interpretation of addition would require that $x + y + z$ be interpreted as $(x + y) + z$. But such a formal interpretation renders such relatively simple manipulations as commuting the two central terms of the polynomial $3x + 2y + 2x + 5y$ into a complex undertaking: $[(3x + 2y) + 2x] + 5y = [3x + (2y + 2x)] + 5y = [3x + (2x + 2y)] + 5y = [(3x + 2x) + 2y] + 5y = (3x + 2x) + (2y + 5y)$. Longer renditions of the same problem quickly become intractable.

A typical fix is to treat addition as an n-ary operation (i.e., as leaving the expression $x + y + z$ essentially unparsed [see Drouhard, 1988; Ernest, 1987]). This resolves the problem of complexity, but only at the price of eliminating subtraction as an operation. For without the usual parsing rule, we must represent subtraction as addition of a negative to protect against application of the commutative law for addition in such expressions as $x - y + z$ to produce (the non-equivalent) $x - z + y$. But to consistently represent subtraction as addition of a negative then requires that, say, the *difference of squares* rule be interpreted as the *sum of a square and the negation of a square* rule.

In such cases, we must confront the fact that the rigorous consistency we prize so much in formal mathematical discourse can be fully imposed on students of elementary algebra only at the cost of radical disconnection from the cognitive processes that ultimately root our fluent algebraic manipulation skills. This is part of the cost of tangling with non-representational notions of cognition and learning. For, from a connectionist standpoint:

> Personal rationality ... results from turning the social process of justification inward upon one's own thoughts ... a kind of self-checking. ... Rules, thus, may play an important role as *knowledge that enters into computations*, but this is a fundamentally different role from the one traditionally conceived by philosophers and cognitive scientists, where *rules constitute the computational algorithms themselves*. (Bereiter, 1991, p. 14)

The pedagogical implication is that the Lexical Support System approach sketched above actually is a balancing act between the significant structural elements of algebra that can be articulated to facilitate students' apprehension of algebra as a formal discourse, and minutiae of representation that only can be formalized at the cost of revealing the limitations of such a discourse in the actual conduct of algebraic operations. There is an irony to enculturating students to mathematics as an explicit structural enterprise through an inherently astructural mathematical domain like elementary algebra symbol manipulation. But as Kirshner (2001) concluded (see also Gee's [1992] theory of the social mind):

> Engaging students in rationalizing their internal processes doesn't mean that those processes actually need to be rational. Logical discourse comes from the social function of rationalizing, not from engagement with inherently logical artifacts. (p. 98)

References

Anderson, J. R. (1983). Acquisition of proof skill in geometry. In R.S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach* (pp. 191-219). Palo Alto, CA: Tioga Publishing Company.

Anderson, J. R., & Lebiere, C. (1998a). Introduction. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 1-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., & Lebiere, C. (1998b). Learning. In J. R. Anderson & C. Lebiere (Eds.), *The atomic components of thought* (pp. 101-142). Mahwah, NJ: Lawrence Erlbaum Associates.

Awtry, T. (1993). *Visual transformations in symbolic elementary algebra*. Unpublished doctoral dissertation, Louisiana State University.

Awtry, T., & Kirshner, D. (1994). Visual salience in algebraic transformations. In D. Kirshner (Ed.), *Proceedings of the Sixteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 131-137). Baton Rouge, LA: Louisiana State University.

Bechtel, W., & Abrahamsen, A. (2002). *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks* (2nd ed.). Malden, MA: Blackwell.

Bell, E. T. (1936). The meaning of mathematics. In W. D. Reeve (Ed.), *The place of mathematics in modern education* (pp. 136-181). Eleventh Yearbook of the National Council of Teachers of Mathematics. New York: Bureau of Publications, Teachers College, Columbia University.

Bereiter, C. (1991). Implications of connectionism for thinking about rules. *Educational Researcher*, *20*, 10-16.

Booth, L. R. (1984). *Algebra: Children's strategies and errors*. Windsor, Berkshire, UK: NFER-NELSON Publishing Company.

Booth, L. R. (1989). A question of structure. In C. Kieran & S. Wagner (Eds.), *Research agenda for mathematics education: Research issues in the learning and teaching of algebra* (pp. 57-59). Reston, VA: National Council of Teachers of Mathematics; Hillsdale, NJ: Lawrence Erlbaum Associates.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47,* 139-159.

Bundy, A., & Welham, B. (1981). Using meta-level inference for selective application of multiple rewrite rule sets in algebraic manipulation. *Artificial Intelligence, 16*, 189-212.

Cajori, F. (1924). *A history of elementary mathematics: With hints on methods of teaching*. New York: The Macmillan Company.

Carry, L. R., Lewis, C., & Bernard, J. E. (1980). *Psychology of equation solving: An information processing study* (Tech. Rep. SED 78-22293). Austin: The University of Texas at Austin, Department of Curriculum and Instruction.

Clancey, W. J. (1999). *Conceptual coordination: How the mind orders experience in time*. Mahwah, NJ.: Lawrence Erlbaum Associates.

College Entrance Examination Board. (1959). *Report of the Commission on Mathematics*. New York: Author.

Cuoco, A., Goldenberg, E. P, & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *Journal of Mathematical Behavior*, *15*(4), 375-402.

Davis, R. B. (1979, April). *Error analysis in high school mathematics, conceived as information-processing pathology*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED171551).

Davis, R. B. (1984). *Learning mathematics*. Norwood, NJ: Ablex.

Davis, R. B., & McKnight, C. C. (1979). Modeling the processes of mathematical thinking. *Journal of Children's Mathematical Behavior*, *2*(2), 91-113.

Dreyfus, H. L. (2002). Intelligence without representation — Merleau-Ponty's critique of mental representation: The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences, 1*(4), 367-383.

Drouhard, J. (1988, July). *Syntactic analysis of algebra and teaching*. Paper presented at the Sixth International Congress on Mathematics Education, Budapest, Hungary.

Dupuy, J-P. (2000). *The mechanization of the mind: On the origins of cognitive science*. (M. B. DeBevoise, Trans.). Princeton, NJ: New French Thought.

Erlwanger, S. H. (1973). Benny's conception of rules and answers in IPI mathematics. *Journal of Mathematical Behavior*, *1*(2), 7-26.

Ernest, P. (1985). The philosophy of mathematics and mathematics education. *International Journal of Mathematics, Education, Science, and Technology, 16*(5), 603-612.

Ernest, P. (1987). A model of the cognitive meaning of mathematical expressions. *British Journal of Educational Psychology, 57*, 343-370.

Ernest, P. (1998). *Social constructivism as a philosophy of mathematics*. Albany, NY: State University of New York Press.

Estep, M. (2003). *A theory of immediate awareness: Self-organization and adaptation in natural intelligence.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Evans, J. St. B. T. (1982). *The psychology of deductive reasoning*. London: Routledge & Kegan Paul.

Fey, J. T. (1992). *Computer-intensive algebra: An overview of fundamental mathematical and instructional themes*. Unpublished manuscript.

Fey, J. T., & Graeber, A. (2003) From the new math to the *Agenda for Action.* In G. Stanic & J. Kilpatrick (Eds.) *History of school mathematics* (Vol. 1, pp. 521-558). Reston, VA: National Council of Teachers of Mathematics.

Freeman, W. J. (1991). The physiology of perception. *Scientific American, 264,* 78-85.

Gee, J. P. (1992). *The social mind: Language, ideology, and social practice*. New York: Bergin & Garvey.

Greeno, J. G. (1982, April). *A cognitive learning analysis in algebra*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Haag, V. H. (1961). *Studies in mathematics*: *Vol. 3. Structure of elementary algebra* (Rev. ed.). New Haven, CT: School Mathematics Study Group, Yale University.

Haberlandt, K. (1997). *Cognitive psychology*. Boston: Allyn & Bacon.

Hamilton, W. R. (1836). A review of Rev. William Whewell's *Thoughts on the Study of Mathematics as a part of Liberal Education*. *Edinburgh Review, 62,* 218-252.

Hamilton, W. R. (1837). Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time. *Transactions of the Royal Irish Academy, XVII*, 293-422.

Haugland, J. (1985). Semantic engines: An introduction to mind design. In J. Haugland (Ed.), *Mind design* (pp. 1-34). Cambridge, MA: : The MIT Press.

Hobson, J. A. (1988). *The dreaming brain*. New York: Basic Books.

Hofstadter, D. R. (1985). Waking up from the Boolean dream, or, subcognition as computation. In D. R. Hofstadter (Ed.), *Metamagical themas: Questing for the essence of mind and pattern* (pp. 635-665). New York: Basic Books.

Kaput, J. J. (1995). A research base supporting long term algebra reform. In D. T. Owens, M. K. Reed, & G. M. Millsaps (Eds.), *Proceedings of the Seventeenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 71-94). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Kirshner, D. (1987). *The grammar of symbolic elementary algebra*. Unpublished doctoral dissertation. University of British Columbia, Vancouver, Canada.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education*, *20*, 274-287.

Kirshner, D. (1998, October) *Discursive algebra: Explicit talk about implicit processes*. Paper and poster presented at the twentieth annual meeting of North American Chapter of the International Group for the Psychology of Mathematics Education, Raleigh, NC.

Kirshner, D. (2001). The structural algebra option revisited. In R. Sutherland, T. Rojano, A. Bell, & R. Lins (Eds.), *Perspectives on school algebra* (pp. 83-98). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Kirshner, D. (2002). Untangling teachers' diverse aspirations for student learning: A crossdisciplinary strategy for relating psychological theory to pedagogical practice. *Journal for Research in Mathematics Education, 33*, 46-58.

Kline, M. (1980). *Mathematics: The loss of certainty*. New York: Oxford University Press.

Lloyd, D. (1989). *Simple minds: Mental representations from the ground up*. Cambridge, MA: MIT Press.

Macintyre, A. (1979/1980). The laws of exponentiation. In A. Dold & B. Eckmann (Proceedings Editors of Model theory and arithmetic, Paris). Published in *Lecture Notes in Mathematics*, *890*, 185-197.

Matz, M. (1980). Towards a computational theory of algebraic competence. *Journal of Mathematical Behavior*, *3*, 93-166.

Merleau-Ponty, M. (1962). *Phenomenology of perception*. London: Routledge & Kegan Paul.

National Advisory Committee on Mathematics Education. (1975). *Overview and analysis of school mathematics grades K-12*. Washington, DC: Conference Board of the Mathematical Sciences.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics, Algebra Working Group. (1998). A framework for constructing a vision of algebra: A discussion document. In National Research Council, *The nature and role of algebra in the K-14 curriculum: Proceedings of a national symposium, May 27 and 28, 1997* (Appendix E, pp. 145-190). Washington, DC: National Academy Press. Available on the Web at http://books.nap.edu/catalog/6286.html.

National Research Council. (1998). *High school mathematics at work: Essays and examples of for the education of all students*. Washington, DC: National Academy Press.

Newell, A., & Simon, H. A. (1985) Computer science as empirical inquiry: Symbols and search. In J. Haugland (Ed.), *Mind design* (pp. 35-66). Cambridge: The MIT Press.

Osborne, A. and Kasten, M. (1992). Change and *An Agenda for Action*: A reconsideration. In R. Morris (Ed.), *Studies in mathematics education: Moving into the twenty-first century* (Vol 8, pp. 21-42). Paris: The United Nations Educational, Scientific and Cultural Organization.

Payne, S. J., & Squibb, H. R. (1990). Algebra mal-rules and cognitive accounts of error. *Cognitive Science*, *14*, 445-481.

Peacock, G. (1833). *Report on recent progress and present state of certain branches of analysis*. In *Report of the Third Meeting of the British Association for the Advancement of Science* (pp. 185-352). London: J. Murray.

Pimm, D. (1995). *Symbols and meanings in school mathematics.* New York: Routledge.

Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, *10*, 163-172.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346-362.

Romberg, T. A., Fennema, E., & Carpenter, T. P. (Eds.). (1993). *Integrating research on the graphical representation of functions*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Royden, H. L. (1968). *Real analysis*. London: Macmillan.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, & J. L., McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 318-362). Cambridge, MA: MIT Press.

Schoenfeld, A. H. (1987). Cognitive science and mathematics education: An overview. In A. H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 1-32). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sfard, A. (1995). The development of algebra: Confronting historical and psychological perspectives. *Journal of Mathematical Behavior, 14*(1), 15-39.

Sleeman, D. (1982). Assessing aspects of competence in basic algebra. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 185-199). New York: Academic Press.

Sleeman, D. (1984). An attempt to understand students' understanding of basic algebra. *Cognitive Science*, *8*, 387-412.

Sleeman, D. (1986). Introductory algebra: A case study of student misconceptions. *Journal of Mathematical Behavior*, *5*(1), 25-52.

St. Julien, J. (1994). *Cognition and learning: The implications of a situated connectionist perspective for theory and practice in education*. Unpublished doctoral dissertation, Louisiana State University.

Thompson, P. W. (1989). Artificial intelligence, advanced technology, and learning and teaching algebra. In C. Kieran & S. Wagner (Eds.), *Research agenda for mathematics education: Research issues in the learning and teaching of algebra* (pp. 135-161). Reston, VA: National Council of Teachers of Mathematics; Hillsdale, NJ: Lawrence Erlbaum Associates.

Thompson, P. W., & Thompson, A. G. (1987). Computer presentations of structure in algebra. In J. Bergeron, N. Herscovics, & C. Kieran (Eds.), *Proceedings of the Eleventh Annual Meeting of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 248-254). Montreal: University of Quebec at Montreal.

Wagner, S., Rachlin, S. L., & Jensen, R. J. (1984). *Algebra learning project: Final report* (Contract No. 400-81-0028). Washington, DC: National Institute of Education.