Understanding of Logical Necessity: Developmental Antecedents and Cognitive Consequences

Anne K. Morris and Vladimir M. Sloutsky

Does abstract reasoning develop naturally, and does instruction contribute to its development? In an attempt to answer these questions, this article specifically focuses on effects of prolonged instruction on the development of abstract deductive reasoning and, more specifically, on the development of understanding of logical necessity. It was hypothesized that instructional emphasis on the metalevel of deduction within a knowledge domain can amplify the development of deductive reasoning both within and across this domain. The article presents 2 studies that examine the development of understanding of logical necessity in algebraic and verbal deductive reasoning. In the first study, algebraic and verbal reasoning tasks were administered to 450 younger and older adolescents selected across different instructional settings in England and in Russia. In the second study, algebraic and verbal reasoning tasks were administered to 287 Russian younger and older adolescents selected across different instructional settings. The results support the hypothesis, indicating that prolonged instruction with an emphasis on the metalevel of algebraic deduction contributes to the development of understanding of logical necessity in both algebraic and verbal deductive reasoning. Findings also suggest that many adolescents do not develop an understanding of logical necessity naturally.

INTRODUCTION

Abstract reasoning allows one to derive conclusions related to classes of objects or events, and it plays an important role in many practical, social, and academic contexts. It has long been debated whether formal instruction contributes to the development of abstract reasoning and whether abstract reasoning is a domain-specific or a domain-general competence (see, e.g., Holyoak & Nisbett, 1988; Johnson-Laird & Byrne, 1991; Lehman, Lempert, & Nisbett, 1988; Overton, 1990; Rips, 1994; Singley & Anderson, 1989, for reviews).

This article examines one facet of this problem. We specifically consider long-term effects of prolonged formal instruction in one domain of knowledge, on the development of abstract reasoning, both within and across this domain. Furthermore, we limit ourselves to the effects of prolonged instruction on the development of understanding of the logical necessity of deductively derived conclusions. Does prolonged domain-specific instruction focused on abstract principles of reasoning have long-term effects on understanding of logical necessity, both within and outside this domain? And do the majority of children develop an understanding of logical necessity without such instruction?

Two studies presented here examine the development of understanding of logical necessity in algebraic and verbal deductive reasoning. The results indicate that understanding of logical necessity does not develop naturally, but that its development can be greatly amplified by prolonged formal instruction.

The Development of Understanding of Logical Necessity

Since Bacon, there has been a distinction between deductive and inductive inference. In inductive inference, the reasoner reaches a tentative conclusion based on available empirical information. Inductive conclusions are probabilified by the premises and may change when new information becomes available—for example: French speak French, Italians speak Italian, Russians speak Russian, Koreans? . . . speak Korean, Americans . . . ?

In contrast, in deductive inference, the reasoner can, based solely on information given in the premises, reach a conclusion that is certain. Those conclusions that are derived in accordance with the rules of deductive inference are said to be necessitated by the premises. These necessitated conclusions are defined as logically necessary, and arguments whose conclusions are necessitated by the premises are defined as valid deductive arguments. Deductive arguments warrant the transition from true premises to true conclusions; therefore, the logically necessary conclusion derived from true premises is true by necessity and a priori. That is, such a conclusion is certain to be true as long as the premises are true and its truth status does not require empirical examinations.

1. We are grateful to the logician Neil Tennant, who suggested this term to us.

© 1998 by the Society for Research in Child Development, Inc. All rights reserved. 0009-3920/98/6903-0095$01.00
Consider the following argument. All A are B. All B are C. Therefore, all A are C. If the premises are instantiated by empirically true statements, for example, All elephants are mammals and All mammals are animals, then the conclusion All elephants are animals is a logically necessary conclusion that is true by necessity and a priori. Understanding that when a conclusion is deduced from true premises, it is true by necessity and is true a priori, is defined as understanding of logical necessity, an ability that is critical for generating and processing deductive arguments.

Understanding of logical necessity relates to the metalevel of reasoning, rather than the transformational level. The distinction between the two levels has been discussed in the literature (Braine, 1990; Johnson-Laird & Byrne, 1991; Moshman, 1990; Polk & Newell, 1995; Sternberg, 1982, 1984). The transformational level is defined as the ability to derive a conclusion, whereas the metalevel is defined as the ability to justify the derived conclusions and to distinguish between probabilified and necessitated conclusions. For example, the ability to perform algebraic transformations on the statement \( a - c + d = d + b - c \), and to derive that \( a = b \), requires the development of transformational level abilities. On the other hand, understanding that the conclusion is logically necessary and that the relation between \( a \) and \( b \) holds for every real number, requires the development of metalevel abilities of deductive reasoning. Because metalevel abilities require abstraction from content-specific propositions, it is likely that these abilities are domain-general and, therefore, transferable from one domain of knowledge to other domains.

The development of metalevel abilities requires the development of transformational level abilities; therefore, the former and the latter are likely to be in hierarchical relations. It has been established that metalevel reasoning abilities develop ontogenetically later than transformational level abilities. Findings indicate that, typically, preschool and elementary school children are capable of formulating some deductive arguments (Hawkins, Pea, Glick, & Scribner, 1984; Komatsu & Galotti, 1986; Kuhn, 1977; Miller, Seier, & Nassau, 1995; Moshman, 1990; Osherson & Markman, 1975); however, they do not distinguish between logically necessary conclusions and conclusions of other types (Komatsu & Galotti, 1986; Miller et al., 1995; Osherson & Markman, 1975; Piaget, 1986). Some researchers have contended that understanding of logical necessity is typically reached by the stage of formal operations (Piaget, 1987; Smith, 1993).

There is little disagreement among researchers with respect to the lack of understanding of logical necessity by preschool and elementary school children; there is much more disagreement regarding adolescents’ and adults’ abilities to understand logical necessity. There is a growing body of evidence that adolescents and adults commit multiple reasoning errors, indicating that reasoners either do not understand logical necessity or fail to apply their understanding of logical necessity to specific reasoning tasks (see, e.g., Evans, 1989; Johnson-Laird & Byrne, 1991, for discussions).

For example, it has been shown that children, adolescents, and adults are prone to belief biases in their deductive reasoning (Evans, 1989; Evans & Pollard, 1990; Markovits & Nantel, 1989; Oakhill & Johnson-Laird, 1985; Oakhill, Johnson-Laird, & Garnham, 1989; Sloutsky, Morris, & Eynon, 1995; see also Evans, Newstead, & Byrne, 1993, for a review). They are more likely to evaluate a deductive argument as valid when it has a more probable, more believable conclusion (“Some religious people are not priests”) than when it has an empirically less probable, less believable conclusion (“Some priests are not religious people”).

It has also been shown that adolescent and adult reasoners often confuse necessitated and probabilized conclusions. For example, many adolescents interpret valid deductive arguments, including valid mathematical proofs, as nonconclusive evidence supporting a claim, or they interpret evidence (e.g., a large number of supporting instances) as conclusive proof (Balacheff, 1988; Bell, 1976; Chazan, 1993; Fischbein & Kedem, 1982; Lee & Wheeler, 1987, 1989; Martin & Harel, 1989; Forcous, 1986; Williams, 1979).

There is also a lack of agreement with respect to the factors contributing to the development of understanding of logical necessity. First, a number of researchers (Inhelder & Piaget, 1958; Overton, 1990) have argued that abstract reasoning develops naturally and that formal instruction does not play a significant role in its development. Second, other researchers contend that some aspects of reasoning may develop naturally but that instruction is an important factor contributing to improvement in abstract reasoning (Cheng, Holyoak, Nisbett, & Oliver, 1986; Fong, Krantz, & Nisbett, 1986; Lehman et al., 1988; Lehman & Nisbett, 1990; Nisbett, Fong, Lehman, & Cheng, 1987). Third, there is a position asserting that abstract reasoning does not develop naturally and that instruction is a necessary condition for its development (Luria, 1976; Vygotsky, 1962).
Theoretically, we side with the position that instruction is a necessary condition for the development of abstract reasoning (Luria, 1976; Vygotsky, 1962). However, it is impossible to provide conclusive evidence for such a strongly formulated position within a single study. Rather, it will require an accumulation of supporting evidence across a number of studies.

This theoretical position has been partially supported by a number of studies conducted in remote societies (Luria, 1976; Scribner & Cole, 1981) and by training studies conducted in modern technological societies (Cheng et al., 1986; Fong et al., 1986; Lehman et al., 1988; Lehman & Nisbett, 1990; Nisbett et al., 1987). The studies in the first set indicate that, for members of remote cultures, instruction is a necessary condition for the development of abstract reasoning, whereas the second set of studies have revealed that for members of modern technological societies, certain types of instruction facilitate the participants’ abstract reasoning. However, the combined findings do not allow us to determine whether or not instruction is a necessary, or merely facilitative, factor in the development of abstract reasoning among members of modern societies.

There is also evidence that instruction focused on the metalevel of deduction (e.g., inferential validity or pragmatic inference schemata) dramatically increased adolescents’ and young adults’ performance on tasks that required understanding of logical necessity (Cheng & Holyoak, 1985; Moshman & Franks, 1986; see also Cheng et al., 1986). Therefore, it seems plausible that prolonged formal instruction that focuses on the metalevel of reasoning and introduces abstract metalevel concepts and relations (e.g., proof, validity, and so on) and symbolism necessary for representing these concepts and relations, is capable of amplifying the development of metalevel abilities, including that of understanding of logical necessity. These considerations led us to the advancement of the following specific hypothesis: Students who receive instruction with a specific emphasis on the metalevel of deduction in a specific knowledge domain (e.g., mathematics) will exhibit a better understanding of logical necessity and a greater developmental progression in understanding of logical necessity, both within and outside this domain, than students who do not receive such instruction.

**Settings**

Because it was not feasible to provide training for a significant number of students for a prolonged period of time, natural settings were identified where students had experienced such instruction. It was clear from the outset that such a design had a number of limitations, including potential confounds. On the other hand, the fact that students received training and were tested within their natural environments had the potential of increasing the ecological validity of the study.

Two settings that provided prolonged instruction in mathematics that focused on either the development of metalevel abilities or transformational level abilities of mathematical reasoning were selected for this study. The first instructional setting was an "experimental" elementary mathematics curriculum in Russia (Davydov, 1975, 1990). This curriculum emphasizes the development of the metalevel of algebraic reasoning, particularly the ability to distinguish between logically necessary conclusions and empirical conclusions (Davydov, 1990). It also emphasizes the logical necessity of particular (e.g., numerical) conclusions that are derived from general mathematical principles and relations, where those principles and relations are first expressed algebraically.

Concepts of quantity, relations, and mathematical structure are emphasized prior to computational or algebraic transformations. The instruction is based on a top-down approach (Vygotsky, 1962) to acquisition of mathematical concepts that starts with a definition followed by instantiations. For example, the teacher introduces the notion of quantity, which is instantiated by physical objects (e.g., large and small containers with water, long and short rulers, and so forth). Then the teacher explains that two objects have some quantitative property if one of the following relations can be applied to the property in question: "equal to," "more than," or "less than." After that, the teacher explains properties of these relations, such as transitivity. Then a problem involving transitivity is posed: "Line A is longer than line B, and line B is longer than line C (lines A and B are drawn on the board). We do not have line C here. Will it be shorter than A, longer than A, or equal to A, given these conditions?" (Davydov, 1975). Students are exposed to this curriculum for 5 elementary school years. This curriculum lays the groundwork for the acquisition of transformational components during a traditional middle school algebra course.

The second "experimental" instructional setting has a different theoretical orientation. National Mathematics Project (NMP) in England (Harper et al., 1987) has an explicit emphasis on the development of transformational components of algebraic reasoning. It tends to replicate the natural progression in
the development of reasoning—the progression from more concrete transformational components to more abstract metacomponents, from computation to understanding of abstract algebraic principles. Transformational components are taught in an inductive case-based manner via the investigation of a number of particular instances. The validity of algebraic generalizations is assessed via empirical checks. For example, the following task is typical within this curriculum: “Find the number that the letter replaces in the equation; guess and check until you find a correct answer: $3(n + 4 = 16)$.” In finding these numerical replacements, students are advised to continue checking until they are “certain” that they have found all possible correct replacements. (For a more comprehensive discussion of both curricula, see Morris, 1995).

Both curricula were implemented in nonselective public schools, serving predominantly middle-class populations of students. These curricula were introduced for whole schools; thus no selection criteria were set for admitting students to these curricular settings. Because these curricula were implemented in different cultures, instructional variables were confounded with other potentially important variables such as language, family processes, and cultural beliefs and practices—variables that may affect academic performance (Chapman, Skinner, & Baltes, 1990; Hess et al., 1986; Okagaki & Sternberg, 1993; see also Murphey, 1992, for a review). To control potential cross-cultural confounds and to allow within-cultural as well as cross-cultural comparisons, schools using “nonexperimental” curricula were selected in the same countries. These schools were located in the same geographical area and had comparable student and teacher populations; however, they did not have curricula that were designed to develop specific kinds of algebraic reasoning. The nonexperimental curriculum in Russia does not specifically focus on metalevel abilities, but it teaches algebraic reasoning at a relatively abstract level. The algebra course for grades 7–9 is “characterized by the enhancement of the theoretical level of instruction and by the stronger emphasis that is gradually placed upon the role of theoretical generalizations and deductions” (USSR Academy of Pedagogical Sciences’ Scientific Research Institute of Curriculum and Teaching Methods, 1987, p. 67). The English nonexperimental curricula emphasize the ability to perform algebraic transformations that are taught in an inductive manner; generalizations are derived from an examination of particular cases. This curriculum tends to reduce the emphasis on formal algebra by using more numerical problem solving.

STUDY 1

Method

Design

The study has a cross-sectional/cross-cultural design that includes two explanatory variables, (1) group (instructional setting nested within culture) and (2) age; and one outcome variable—understanding of logical necessity. To distinguish understanding of logical necessity from transformational level abilities, measures of the ability to perform domain-specific transformations were also included in the design. To control some potential extraneous effects, background information regarding academic coursework was collected in each setting.

Participants

Four groups were included in the sample: (1) Russian experimental students and graduates of an experimental elementary school mathematics curriculum, implemented in a school in Moscow ($n = 120$); (2) Russian nonexperimental students in a nonexperimental school in Moscow ($n = 89$); (3) English experimental students in an upper school in England that implemented the National Mathematics Project curriculum ($n = 120$); and (4) English nonexperimental students in an upper school in England with a “nonexperimental” curriculum ($n = 120$). Each group included students at two age levels—early adolescents (“younger group”) and middle and late adolescents (“older group”). For all groups, students had not had explicit instruction in formal logic. Details of the sample and students’ coursework are provided in Table 1 and Table 2. Data in Table 2 suggest that younger students in the Russian groups had more extensive exposure to mathematics and native language studies than younger students in the English groups, whereas older students across all groups had comparable exposure to major academic subjects.

These age groups were selected on the basis of their exposure to formal algebra. We required that students in the younger group, under each condition, would have at least 1 year of exposure to formal algebra and that students in the older group would have at least 3 years of formal algebra instruction. Our attempt to match students’ exposure to algebra resulted in the following: for the younger group, students in the Russian samples were more than a year younger than their English peers (see Table 1). This was not considered to be a problem for the study, however, because the direction of performance dif-
Table 1  Detailed Description of the Sample: Total Number of Participants, Gender, Mean Ages, and Standard Deviations of the Means by Groups

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>N of Boys</th>
<th>N of Girls</th>
<th>Mean Age</th>
<th>SD of Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>62</td>
<td>36</td>
<td>26</td>
<td>12.1</td>
<td>.77</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>55</td>
<td>26</td>
<td>29</td>
<td>12.3</td>
<td>.55</td>
</tr>
<tr>
<td>English experimental</td>
<td>32</td>
<td>14</td>
<td>18</td>
<td>13.9</td>
<td>.26</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>35</td>
<td>16</td>
<td>19</td>
<td>13.8</td>
<td>.19</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>59</td>
<td>32</td>
<td>27</td>
<td>15.45</td>
<td>.59</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>34</td>
<td>14</td>
<td>20</td>
<td>15.2</td>
<td>.65</td>
</tr>
<tr>
<td>English experimental</td>
<td>88</td>
<td>42</td>
<td>46</td>
<td>15.36</td>
<td>.64</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>85</td>
<td>40</td>
<td>45</td>
<td>15.39</td>
<td>.61</td>
</tr>
</tbody>
</table>

Differences was hypothesized to be opposite to the direction of the age differences.

For the purposes of sample selection, each school was contacted by a member of a research group. After permission to conduct the study was granted, in each school, two intact classes in each grade level were selected. From this pool, students were then randomly selected for inclusion in the study and contacted by the researchers. All contacted students agreed to participate in the study. Sample selection in Russia was arranged and overseen by the authors with the assistance of the Russian Academy of Education. Sample selection in England was carried out by one of the authors.

Materials

Instruments were designed to compare students’ performance on tasks that required understanding of logical necessity. To control the overall level of transformational proficiency, reasoning tasks that required only transformational abilities were also included in the study. All instruments were prepared in English and then translated into Russian. After a back translation, necessary revisions were made in the Russian equivalents of the tasks. The following tasks were used in this study.

Algebraic Reasoning Tasks

Tasks that require understanding of logical necessity.

1. Victor Problem (modification of Fischbein & Kedem, 1982).

In an algebra class, the teacher proved that every whole number of the form \( n^3 - n \) is divisible by 6 with no remainder. The proof was as follows:

We can write:

\[ n^3 - n = n(n^2 - 1). \]

Then we can rewrite the expression on the right:

\[ n(n^2 - 1) = n(n - 1)(n + 1). \]

So:

\[ n^3 - n = n(n - 1)(n + 1) = (n - 1)(n)(n + 1). \]

But \((n - 1)(n)(n + 1)\) is a product of three

Table 2  Coursework in the Studied Settings: Hours per Year in Major Academic Domains

<table>
<thead>
<tr>
<th>Groups</th>
<th>Math</th>
<th>Native Language</th>
<th>Foreign Language</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>192</td>
<td>320</td>
<td>128</td>
<td>160</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>192</td>
<td>256</td>
<td>64</td>
<td>160</td>
</tr>
<tr>
<td>English experimental</td>
<td>114</td>
<td>114</td>
<td>76</td>
<td>152</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>133</td>
<td>158</td>
<td>133</td>
<td>190</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>160</td>
<td>192</td>
<td>128</td>
<td>160</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>160</td>
<td>192</td>
<td>106</td>
<td>160</td>
</tr>
<tr>
<td>English experimental</td>
<td>114</td>
<td>133</td>
<td>95</td>
<td>152</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>133</td>
<td>158</td>
<td>95</td>
<td>190</td>
</tr>
</tbody>
</table>
consecutive whole numbers. Therefore one of them should be divisible by 2, and one of them (not necessarily a different one) should be divisible by 3. Thus their product should be divisible by \(2 \times 3\), that is, by 6.

Victor [Petya, in the Russian version], a member of the algebra class, is a doubter. He thinks that we have to check at least a hundred numbers to be sure that the theorem is correct. What is your opinion? Do you agree with Victor? Explain your answer.

This task, which has been used in previous research (e.g., Fischbein & Kedem, 1982; Vinner, 1983), was intended to measure understanding of proof as a logically valid argument that necessitates the truth of the derived conclusion.


A girl multiplies a number by 5 and then adds 12. She then subtracts the original number and divides the result by 4. She notices that the answer she gets is 3 more than the number she started with. She says, “I think that would happen, whatever number I started with.” Is she right? Prove that your answer is right.

This task, which has been used in previous research (e.g., Lee & Wheeler, 1989), was intended to measure students’ ability (1) to apply their understanding of logical necessity and (2) to formulate a formal proof as a deductive argument.

Tasks that require the ability to perform algebraic transformations. These tasks require a certain level of algebraic proficiency but do not require understanding of logical necessity. The first task measured students’ ability to perform simple algebraic transformations, specifically the ability to write simple equations that establish relations between two quantities. It has been used extensively in previous research with high school and college students (see, e.g., Bernardo & Okagaki, 1994; Clement, Lochhead, & Monk, 1981; Fisher, 1990; Hegarty, Mayer, & Green, 1992; Schoenfeld, 1989). The second task measures the ability to write more complex equations. It has also been used in previous research (Küchemann, 1978).


Write an equation using the letters \(S\) and \(T\) to represent the following statement: “There are six times as many students as teachers in this school.” Use \(S\) for the number of students and \(T\) for the number of teachers.


Mary’s basic wage is 20 pounds per week. She is also paid 2 pounds for each hour of overtime that she works. If \(h\) stands for the number of hours of overtime that she works, and if \(W\) stands for her total wage in pounds, write down an equation connecting \(W\) and \(h\).

Verbal Deductive Reasoning Tasks

Tasks that require understanding of logical necessity. The following task presents a choice between a logically necessary but unbelievable conclusion (26 can be divided evenly by 8) and believable conclusions that do not follow from the premises.

Fahmooth number:

Assume that the first two sentences are true. Make a conclusion from the assumptions. (Choose \(a\), \(b\), \(c\), or \(d\)).

All fahmooth numbers can be divided evenly by 8.

26 is a fahmooth number.

Therefore:

\(a\) 26 must not be a fahmooth number.
\(b\) 26 is an exception to the rule.
\(c\) It is probably true that fahmooth numbers cannot be divided evenly by 8.
\(d\) 26 can be divided evenly by 8.

Tasks that require the ability to perform verbal deductive transformations. This task has the same logical form (categorical syllogism of the first figure and positive-universal form of the premises and the conclusion) as the fahmooth number task. To avoid familiarity effects, the premises and the conclusion were explicitly counterfactual; however, the following instructions urged participants to suspend their empirical beliefs:

Please tell whether or not the sentences show correct reasoning. All the sentences are really nonsense, but you are to think only about the reasoning. Circle YES if the reasoning is good and NO if the reasoning is not good.

Syllogism with explicitly counterfactual information (Wilson, Cahan, & Begle, 1966):

If all birds have purple tails and all cats are birds, then all cats have purple tails.

Procedure

All students were tested in their schools during the first part of the academic year. The tasks were
presented as a written test that was distributed by a research assistant or specially trained regular classroom teacher during the regular math class. The tasks were part of a larger test. Students were given 90 min for the entire test. After the test, students were debriefed by a researcher or research assistant and asked to explain their solutions to the problems. Debriefing interviews took place in a separate room in the school.

Scoring

Data were encoded by two coders in accordance with a coding catalog. Cohen’s kappas, indices of interrater reliability, varied for the described tasks from 0.96 to 1. In response to the Victor problem, students provided three main categories of responses, including: (1) I disagree with Victor because empirical confirmations cannot prove anything or because the algebraic proof is sufficient; (2) I agree with Victor that empirical confirmations are needed (however, 100 confirmations may be too many); and (3) other responses. In response to the Girl problem, students provided three main categories of responses, including (1) formulation of a correct algebraic proof; (2) use of numerical examples; and (3) other responses. In response to the Student-Professor and Mary problems, students provided three main categories of responses, including (1) writing of a correct equation; (2) writing of an incorrect equation; and (3) other responses. In verbal reasoning tasks, the answer was scored YES if the student provided a correct answer and NO if the student provided an incorrect answer. If the student did not provide any response, the data were treated as missing.

Results and Discussion

This section will focus on the development of understanding of logical necessity across age groups and instructional settings. Because the preliminary analysis did not reveal significant gender differences, in all subsequent analyses, boys and girls were treated as one group.

For purposes of statistical analyses, the data were cross-tabulated (group \( \times \) age \( \times \) response category). The “group” variable has four levels: (1) Russian experimental group (RE); (2) English experimental group (EE); (3) Russian nonexperimental group (RNE); and (4) English nonexperimental group (ENE). The age variable has two levels: younger adolescents and older adolescents. The response category variable has two levels: (1) yes (when the response fits the category) and (2) no (when the response does not fit the category). Statistical instruments included log-linear and chi-square analyses that allowed us to select appropriate well fitting log-linear models and to determine goodness-of-fit chi-squares, component chi-squares, \( p \) values of the models and of the effects, directions of the effects, and approximations of \( z \) scores of the effects (\( \lambda / \text{SE} \)).

The Development of the Metalevel of Algebraic Reasoning

Frequencies of responses to the Victor and Girl problems by response category, age group, and instructional setting are presented in Tables 3 and 4. In both tasks, older RE students exhibited a better understanding of proof as a means of deriving a logically necessary conclusion than students in other groups. Across the tasks, RE students exhibited marked developmental gains in their understanding of logical necessity within algebraic reasoning.

In the Victor problem, “Agreement with Victor” (and a request for numerical confirmations of a proved conclusion) suggested that the student confused proof as a logically valid argument that necessitates the truth of the derived conclusion with a probabilistic inductive statement that requires additional empirical confirmations. Disagreement with Victor accompanied by an adequate explanation (understanding that proof is sufficient) was interpreted as understanding of the logical necessity of the proved conclusion.

In the Girl problem, the use of a correct algebraic proof as a solution to the problem suggested that the student (1) understands that only a valid proof, as an argument that necessitates the conclusion, can prove that the girl is right (whereas numerical examples can only exemplify it) and (2) can apply an understanding of the logical necessity of proof to the solution of a specific task. Use of numerical examples not accompanied by proof suggested that the student confuses evidence with proof and therefore does not understand logical necessity or cannot apply this understanding in the solution of specific reasoning tasks.

In both the Victor problem (response category “Proof is sufficient”) and the Girl problem (response category “Formulated correct algebraic proof”), models with effects due to group and age fit the data well. In the Victor problem (see the second column in Table 3), the direction of the effects suggested that older students in the RE group were more likely to treat the proof as a valid argument than students in other groups, \( z = 4.5, p < .001 \). The data also indicate (see the third column in Table 3) that students in the English groups requested numerical confirmations of
Table 3 Percentages of Responses in Each Response Category and Examples of Responses to the Victor Problem

<table>
<thead>
<tr>
<th>Groups</th>
<th>Proof Is Sufficient</th>
<th>Empirical Confirmations Are Needed</th>
<th>Other Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>3</td>
<td>8</td>
<td>89</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>0</td>
<td>13</td>
<td>87</td>
</tr>
<tr>
<td>English experimental</td>
<td>3</td>
<td>63</td>
<td>34</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>0</td>
<td>54</td>
<td>46</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>51</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>9</td>
<td>18</td>
<td>73</td>
</tr>
<tr>
<td>English experimental</td>
<td>7</td>
<td>68</td>
<td>25</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>6</td>
<td>49</td>
<td>45</td>
</tr>
<tr>
<td>Fitted log-linear model</td>
<td>Group + Age, $\chi^2(3, N = 450) = 4.63$, $p = .2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significance of effects</td>
<td>Group, Component $\chi^2(3) = 50$, $p &lt; .001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age, Component $\chi^2(1) = 41$, $p &lt; .001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examples of answers falling into each category</td>
<td>I think that Victor is right to be doubtful, just because it works for some numbers does not mean that it works for others.</td>
<td>(1) I did not understand the proof. (2) Peta has an inferiority complex—he needs to see a psychotherapist.</td>
<td></td>
</tr>
</tbody>
</table>

the proved statement more frequently than students in the Russian groups, indicating a confusion of proof with empirical evidence, $\chi^2(1, N = 184) = 26, p < .001$, for the younger group, and $\chi^2(1, N = 266) = 17, p < .001$, for the older group.

In the Girl problem (see second column in Table 4), the direction of the effects suggested that younger students in the RE group were more likely to use algebraic proof than younger students in the other groups, $z = 2.24, p < .01$, whereas older students in both Russian groups were more likely to use algebraic proof than students in the English groups, $z = 6.79, p < .001$.

The developmental progression in understanding

Table 4 Percentages of Responses in Each Response Category in the Girl Problem

<table>
<thead>
<tr>
<th>Groups</th>
<th>Formulated Correct Algebraic Proof</th>
<th>Used Numerical Examples Only</th>
<th>Other Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>19</td>
<td>26</td>
<td>55</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>11</td>
<td>31</td>
<td>58</td>
</tr>
<tr>
<td>English experimental</td>
<td>0</td>
<td>84</td>
<td>16</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>0</td>
<td>57</td>
<td>43</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>69</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>41</td>
<td>35</td>
<td>24</td>
</tr>
<tr>
<td>English experimental</td>
<td>9</td>
<td>73</td>
<td>18</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>4</td>
<td>42</td>
<td>54</td>
</tr>
<tr>
<td>Fitted log-linear model</td>
<td>Group + Age, $\chi^2(3, N = 450) = .85$, $p = .84$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significance of effects</td>
<td>Group, Component $\chi^2(3) = 115$, $p &lt; .001$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Age, Component $\chi^2(1) = 45$, $p &lt; .001$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of logical necessity by task and instructional setting is presented in Figure 1. Data in Figure 1 indicate that across the tasks, students in different groups exhibited differences in their developmental gains in understanding of logical necessity within algebra. Russian experimental students exhibited marked developmental gains in their understanding of logical necessity in the Victor problem, $\chi^2(1, N = 121) = 35.3, p < .001$, and in the Girl problem, $\chi^2(1, N = 121) = 30.9, p < .001$, and the RNE students exhibited a developmental gain in the Girl problem, $\chi^2(1, N = 89) = 11, p < .001$. At the same time, students in both English groups did not exhibit any significant developmental gains in their understanding of logical necessity in algebraic reasoning.

Table 5 presents data on the development of transformational components of algebraic reasoning: the best fitting log-linear models; descriptions of effects; and the significance of group differences in algebraic transformational proficiency. Table 5 shows that for both the Student-Professor and the Mary problems, models with effects due to group and age fit the data reasonably well. The results of the analysis presented in Table 5 indicate that for the Student-Professor problem, older students in the RE group were more likely to write correct equations than students in other groups, whereas younger RE students did not differ significantly from students in the EE group. It should also be noted that all groups, except for the EE group, exhibited significant increases in their performance on this task. No significant developmental gains were found for the EE group.

The results also indicate that for the Mary problem, older students in the RE group formulated correct equations more frequently than students in the other groups, whereas no differences were found for the younger students. In all groups, students exhibited significant developmental gains in their performance on this task.

In sum, the older RE students exhibited not only a better understanding of logical necessity but greater transformational proficiency as well. (These differences, however, were not present for the younger students.) This result suggests a rival explanation for the better metalevel performance of the RE students. That is, the better understanding of logical necessity by the RE students could be attributable to their transformational proficiency. We deemed it necessary to dispute this argument. From the outset, we conceptualized transformational proficiency as a necessary (but not sufficient) condition for the development of metalevel competency. Therefore, high metalevel proficiency should indicate high transformational proficiency, but high transformational proficiency does not necessarily indicate high metalevel proficiency.

To seek support for this, we (1) selected those students who correctly performed on at least one transformational problem and (2) computed how many of these students correctly performed on at least one
Table 5  Percentage of Children Writing Correct Equations, Best Fitting Models, Significant Effects, and Observed Differences in the Student-Professor and Mary Problems

<table>
<thead>
<tr>
<th>Groups</th>
<th>Student-Professor Problem</th>
<th>Mary Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>60</td>
<td>18</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>English experimental</td>
<td>63</td>
<td>13</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>83</td>
<td>75</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>53</td>
<td>38</td>
</tr>
<tr>
<td>English experimental</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>28</td>
<td>19</td>
</tr>
</tbody>
</table>

Fitted log-linear model

\[ \text{Group + Age, } \chi^2(3, N = 450) = 7.4, \quad p = .06 \]

Significance of effects

\[ \text{Group, Component } \chi^2(3) = 78, \quad p < .001 \]

\[ \text{Age, Component } \chi^2(1) = 18, \quad p < .001 \]

Observed differences:

Within the younger group

\[ \text{RE, EE > RNE, ENE, } z = 2.8, \quad p < .001 \]

Within the older group

\[ \text{RE > RNE, EE, ENE, } z = 2.99, \quad p < .001, \text{ RE, RNE > ENE, } z = 3.23, \quad p < .001 \]

The increase from the younger to the older group

\[ \text{RE, } \chi^2(1, N = 121) = 8, p < .01 \]

\[ \text{RNE, } \chi^2(1, N = 89) = 10.4, p < .005 \]

\[ \text{EE, ns} \]

\[ \text{ENE, } \chi^2(1, N = 120) = 14, p < .001 \]

\[ \text{RE, } \chi^2(1, N = 121) = 39, p < .001 \]

\[ \text{RNE, } \chi^2(1, N = 89) = 39, p < .001 \]

\[ \text{EE, } \chi^2(1, N = 120) = 8, p < .01 \]

\[ \text{ENE, } \chi^2(1, N = 120) = 7.6, p < .01 \]

**Note:** Sign > means “more frequently.” For example, RE, EE > RNE, ENE means that students in the Russian and English experimental provided correct responses more frequently than students in the Russian and English nonexperimental groups.

The problem that required understanding of logical necessity (LN/T ratio). We also (3) selected those students who correctly performed on at least one problem that required understanding of logical necessity and (4) computed how many of these students correctly performed on at least one problem that required transformational proficiency (T/LN ratio). Due to the very small number of younger students who exhibited an understanding of logical necessity, these ratios were not computed for the younger groups. As predicted, T/LN ratios were much higher than LN/T ratios. The majority of T/LN ratios were close to 100% (they were 88% in the EE group, 45% in the ENE group, 98% in the RE group, and 87% in the RNE group), whereas LN/T ratios did not surpass 60% (they were 13.5% in the EE group, 17% in the ENE group, 60% in the RE group, and 40% in the RNE group). These data support our theoretical distinction between meta- and transformational level competencies and indicate that differences in understanding of logical necessity cannot be attributed solely to the differences in transformational proficiency.

In sum, students in English groups exhibited minimal understanding of logical necessity and a minimal developmental progression in their understanding of logical necessity. In contrast, many older students in the RE group understood logical necessity (there were no differences for the younger groups), and students in the RE group exhibited a profound developmental progression in their understanding of logical necessity. The results were mixed for the RNE group. The older RNE students performed significantly better than students in English groups, but were significantly less likely to demonstrate an understanding of logical necessity as compared to students in the RE group (there were no such differences for the younger groups); and students in the RNE group exhibited a developmental progression in their understanding of logical necessity. At the same time, in
every group, there were more students exhibiting transformational proficiency than there were students understanding logical necessity.

The Development of the Metalevel of Verbal Reasoning

Our next goal was to examine whether understanding of logical necessity in algebraic reasoning corresponds to that in verbal deductive reasoning. Data pertaining to verbal deductive reasoning are presented in Table 6. The table contains percentages of correct responses, best fitting log-linear models, and description of effects.

Data in Table 6 indicate that younger students in the RE group were less likely to respond correctly to the fathome problem than students in other groups, whereas older students in the RE group performed significantly better on this problem than students in other groups. The results also indicate that students in the RE group exhibited a marked developmental increase in their understanding of logical necessity, whereas students in other groups exhibited no developmental increase.

In the Positive-Universal syllogism, younger RNE students performed significantly better than students in other younger groups. No significant differences were found for the older students. With the exception of the ENE group, students in all groups exhibited significant developmental gains in their ability to perform verbal deductive transformations.

As in algebraic reasoning tasks, in verbal reasoning tasks we also calculated LN/T and T/LN ratios. Across age groups, the LN/T ratios did not surpass 20% (they were 19% for the EE group, 10% for the ENE group, 14.5% for the RE group, and 10% for the RNE group), whereas T/LN ratios did not fall below 78% (they were 96% for the EE group, 85% for the ENE group, 85% for the RE group, and 78% for the RNE group). Therefore, as in algebraic reasoning, it appears unlikely that the observed limitations in the understanding of logical necessity in verbal deductive reasoning stem from transformational limitations. Developmental gains in transformational and

<table>
<thead>
<tr>
<th>Groups</th>
<th>Reasoning Problems</th>
<th>Positive-Universal Syllogism</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>71</td>
</tr>
<tr>
<td>Russian experimental</td>
<td>9</td>
<td>44</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>16</td>
<td>75</td>
</tr>
<tr>
<td>English experimental</td>
<td>14</td>
<td>89</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td>31</td>
<td>86</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>12</td>
<td>74</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>20</td>
<td>92</td>
</tr>
<tr>
<td>English experimental</td>
<td>9</td>
<td>87</td>
</tr>
<tr>
<td>English nonexperimental</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitted log-linear model</td>
<td>Fully saturated</td>
<td></td>
</tr>
</tbody>
</table>

| Significance of effects | Group + Age, $\chi^2(3, N = 450) = 4.07$, $p < .01$
|-------------------------|----------------------------------|
| Group * Age, $\chi^2(3, N = 450) = 13$, $p < .01$
| Age, Component $\chi^2(1) = 13.5$, $p < .001$

<table>
<thead>
<tr>
<th>Observed differences:</th>
<th>Within the younger group</th>
<th>Within the older group</th>
</tr>
</thead>
</table>
| EE, RNE, ENE > RE, $z = 2.13$, $p < .001$ | RE > RNE, EE, ENE, $z = 2.1$, $p < .001$ | ENE > RE, RNE, ENE, $z = 2.1$, $p < .0001$
| RE > RNE, EE, ENE, $z = 3.23$, $p < .001$ |                  |
| The increase from the younger to the older group | RE, $\chi^2(1, N = 121) = 16$, $p < .001$ |
| RNE, $\chi^2(1, N = 89) = 7.6$, $p < .01$ |
| EE, $\chi^2(1, N = 120) = 6.2$, $p < .05$ |

ENE, ns
metalevel abilities of verbal deductive reasoning are presented in Figure 2.

It can be seen from the figure that, similar to algebraic reasoning, in verbal deductive reasoning students in the English groups exhibit no developmental progression in their understanding of logical necessity, whereas students in all groups exhibit a significant developmental progression in their verbal transformational proficiency. The fact that developmental patterns of transformational level abilities differ markedly from developmental changes in metalevel abilities provides some support for the notion that the tasks did tap different levels of deductive reasoning.

These findings point to the possibility that mathematics instruction does affect the development of understanding of logical necessity within verbal deductive reasoning. The results are consistent with those on the development of logical necessity within algebraic reasoning. First, similar to the domain of algebra, younger students in the RE group did not exhibit better understanding of logical necessity than students in other groups, whereas older students in the RE group differed markedly in their understanding of logical necessity from students in other groups. In addition, as in the domain of algebra, students in the RE group exhibited a significant developmental progression in their understanding of logical necessity, whereas students in other groups did not (with the exception of the RNE students on one algebraic task). Finally, as in the domain of algebra, students' limitations in understanding of logical necessity did not seem to stem from their transformational limitations.

The fact that adolescents in both English groups exhibited minimal or no developmental progression in their understanding of logical necessity seems to contradict well-established notions that understanding of logical necessity develops naturally, in the course of normal cognitive development (e.g., Inhelder & Piaget, 1958; Piaget, 1987). Although Inhelder and Piaget (1958) contend that environmental constraints may detrimentally affect the normal course of cognitive development, it does not appear plausible that there were environmental constraints that prevented so many students from developing an understanding of logical necessity. The fact that many students attending regular high schools and living in middle-class families and neighborhoods do not develop an important formal-operational property—understanding of logical necessity—casts doubt on the idea of naturally developing formal operations.

The results partially support the hypothesis of the study, suggesting that mathematics instruction may affect the development of understanding of logical necessity. However, some limitations of the design and the instruments did not allow us to firmly link understanding of logical necessity to instruction. First, the observed differences could be attributed to differences in intelligence among the groups. Second, there are some differences in coursework across the cultures that suggest a potential rival explanation, and the lack of within-group variability in the coursework did not allow us to statistically control this variable. Third, there were many children who reported that they did not understand the proof in
the Victor problem, suggesting that the failure to detect a developmental progression in RNE, EE, and ENE could be due to a floor effect. Fourth, the contribution of transformational level proficiency to performances on metalevel algebraic tasks was controlled across tasks, not within tasks. It is possible, however, that the observed limitations in students’ performances on the Victor and Girl problems could be attributed to the lack of transformational proficiency needed for the processing of these specific tasks. And finally, the instruments allowed us to establish, but not to quantify similarities between the development of logical necessity in algebraic and verbal deductive reasoning. Therefore, we deemed it necessary to conduct a second study that would take into consideration these limitations. The goal of the second study was to refine the instruments and to introduce tighter control to provide more definitive evidence pertaining to the effects of instruction on the development of understanding of logical necessity.

STUDY 2

The study was designed to test the original hypothesis, while controlling potential confounds. There are three aspects in which this study differs from the first study. First, to match samples with respect to students’ coursework, this study was limited to Russian students. Second, instruments were substantially revised. In particular, the proof in the Victor problem was simplified, and an attempt was made to introduce within-task controls of algebraic transformational proficiency. To do so, we attempted to differentiate students’ ability to represent quantitative relations with an equation, from their certainty that the conclusion derived from that equation must be universally valid. Subjective certainty, as a measure of understanding of logical necessity, has been used in the past (Miller, 1986). The application of this measure, however, sparked criticism (Foltz, Overton, & Ricco, 1995; Murray, 1990; Overton, 1990). Critics contend that the fact that the reasoner has a high degree of certainty about the truth of a conclusion does not necessarily demonstrate that this reasoner understands that the conclusion is logically necessary. For example, one can have a high degree of certainty that a conclusion is true because it has always held in the past. Although it is difficult to disagree that a high degree of subjective certainty does not necessarily mean that the reasoner does understand logical necessity, it seems obvious that a low degree of subjective certainty in deductively derived conclusions signifies a lack of understanding of logical necessity. Thus subjective certainty can be regarded as a necessary, but not sufficient, condition for concluding that the reasoner does understand logical necessity. To infer that the reasoner did indeed understand logical necessity, we deemed it important to use this measure in conjunction with other measures. Finally, to eliminate a potential rival explanation, measures of students' intelligence were introduced.

Method

Design

The study had a cross-sectional design that included instructional setting and age as explanatory variables, and understanding of logical necessity in algebraic and verbal deductive reasoning as outcome variables.

Participants

Two groups of students (who did not participate in Study 1) were included in the sample: (1) Russian experimental (RE)—students and graduates of the same experimental elementary school mathematics curriculum as that in Study 1, implemented in a school in Moscow (n = 135); and (2) Russian nonexperimental (RNE)—students in a nonexperimental school in Moscow (n = 152). Each group included students at two age levels—early adolescents (“younger group”) and middle and late adolescents (“older group”). Both schools were located in the same area of the city, drawing from predominately middle-class neighborhoods. These schools did not differ in the amount of coursework. After their experiences in the experimental elementary curriculum, students in the RE group experienced the same middle and high school curriculum as RNE students.

To control for intelligence, measures of the participants’ intelligence were included. The Russian version of WISC-R (Pansiyuk, 1973) was administered to the participants. To make intelligence scores comparable across the age groups, the raw scores for each subtest were standardized across the whole sample with $M = 10$ and $SD = 3$. These standardized scores summed across the subtests and averaged for each subgroup are shown with other descriptive statistics in Table 7.

Materials

Algebraic Reasoning Task

These tasks were designed to examine the adolescents’ ability to appreciate the logical necessity of a deductively derived conclusion.
Table 7 Descriptive Statistics of the Sample

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>n of Boys</th>
<th>n of Girls</th>
<th>Mean Age</th>
<th>SD of Age</th>
<th>Mean IQ</th>
<th>SD of IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Younger group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>82</td>
<td>44</td>
<td>38</td>
<td>12.6</td>
<td>1.1</td>
<td>108</td>
<td>3.6</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>93</td>
<td>31</td>
<td>62</td>
<td>12.6</td>
<td>1.1</td>
<td>109</td>
<td>5.5</td>
</tr>
<tr>
<td>Older group:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian experimental</td>
<td>53</td>
<td>34</td>
<td>19</td>
<td>15.6</td>
<td>.5</td>
<td>111</td>
<td>3.9</td>
</tr>
<tr>
<td>Russian nonexperimental</td>
<td>59</td>
<td>20</td>
<td>39</td>
<td>15.6</td>
<td>.5</td>
<td>112</td>
<td>4.3</td>
</tr>
</tbody>
</table>

1. Girl problem (revised). The problem itself was worded in exactly the same way as that in Study 1. However, the questions posed to the participant were different. The participants were asked (1) to solve the problem proving that the girl is right; (2) to indicate whether or not they would need to include numerical confirmations as part of a proof to prove to other people that the girl is right (and if so, how many numerical confirmations should be included in a proof); and (3) to evaluate their certainty that the girl’s conclusion is correct. In part 3, the participants rated their certainty using a 4 point Likert-type scale that ranged from 0 (no certainty) to 3 (absolutely certain) under two imaginary conditions: (1) a proof was presented, but no numerical confirmations were carried out, and (2) no proof was presented, but the regularity was confirmed with more than 100 numbers. Part 1 examined students’ algebraic proficiency, whereas parts 2 and 3 examined students’ understanding of logical necessity.

2. Victor problem (revised). As compared to Study 1, the proof in this problem was substantially simplified and new questions were added to the task.

In an algebra class the teacher proved that for every natural number \( n \), the expression \( n^2 + n \) is even (that is, if you divide \( n^2 + n \) by 2, there will be no remainder). The proof was as follows:

It is true for every natural number \( n \) that:

\[ n^2 + n = n(n + 1). \]

But \( n(n + 1) \) is a product of two consecutive natural numbers. That is, we are just multiplying two consecutive natural numbers.

Therefore, one of the numbers should be divisible by 2. That is, either \( n \) or \( n + 1 \) can be divided by 2 with no remainder. Thus their product is divisible by 2.

Therefore, because \( n(n + 1) \) is divisible by 2, it is even. Therefore, \( n^2 + n \) is even.

Victor is a member of the algebra class. Victor understood every step in the teacher’s proof that for every natural number \( n \), \( n^2 + n \) is even. However, Victor is a doubter. He thinks that we have to test at least a hundred numbers to be sure that the rule is correct.

Participants were asked (1) whether or not they understood the proof; (2) whether or not they agreed with Victor; (3) why they agreed or disagreed with Victor; and (4) to evaluate their certainty that the rule worked for all natural numbers. In part 4, participants rated their certainty using a 4 point Likert-type scale that ranged from 0 (no certainty) to 3 (absolutely certain) under two imaginary conditions: (1) the proof was presented but no numerical confirmations were carried out, and (2) no proof was presented but the regularity was confirmed with more than 100 numbers.

Part 1 was designed to examine students’ algebraic proficiency, whereas parts 2, 3, and 4 examined students’ understanding of logical necessity.

Verbal Deductive Reasoning Task

Because transformational competency was controlled within algebraic reasoning tasks, all verbal deductive reasoning tasks were designed to examine students’ understanding of logical necessity within verbal deductive reasoning. Thirteen categorical syllogisms were included in the study. All syllogisms contained counterfactual content and had the following logical form:

\[ \text{All } A \text{ are } B. \]
\[ \text{A ? is an } A. \]

Instructions to these syllogisms explicitly asked students to suspend their beliefs about information in the premises and in the conclusion, to assume that all premises are true, and to pay attention only to the logical form of the syllogisms. In four of these syllogisms, students were presented with possible conclusions. An example of one of these syllogisms is shown below (the asterisk identifies the conclusion that logically follows from the premises):
All felismo animals are not predators.
The lion is a felismo animal.
Therefore:
1. The lion must not be a felismo animal.
2. The lion is an exception to the rule.
3.* The lion is not a predator.
4. It is probably not true that all felismo animals
are not predators.

To eliminate potential transformational limitations, in the other nine syllogisms, students were presented with one conclusion and they were asked if the conclusion logically followed from the premises. Two examples below represent such syllogisms:

A. Conclusion follows from the premises (valid argument).
   All compounds that have gnocological ele-
ment are flammable.
   Water has a gnocological element.
   Therefore water is flammable.
B. Conclusion does not follow from the premises
   (nonvalid argument).
   All mammals have a highly organized brain.
   The elephant’s brain is highly organized.
   Therefore, the elephant is a mammal.

Conclusions to these arguments varied in their believeability (the believeability of conclusions was measured in a separate study conducted with the same students after the completion of the present study). A full list of syllogisms and measures of believeability of conclusions is presented in the Appendix. In these syllogisms, students were expected to take into consideration the logical form of the argument rather than the believeability of its conclusion.

Scoring

In the Girl problem, the following components of students’ responses were scored: (1) use of correct equations (1 if they used correct equations, otherwise 0); (2) the necessity of numerical confirmations of proved statements (1 if they did not require confirmations, otherwise 0); and (3) subjective certainty in a deductively derived conclusion (1 if they were absolutely certain, otherwise 0). In the Victor problem, the following response categories were scored: (1) understanding of the presented proof (1 if understanding was reported, otherwise 0); (2) disagreement with Victor because the proof is sufficient (1 if disagreement was reported, otherwise 0); and (3) subjective certainty in the deductively derived conclusion (1 if they were absolutely certain, otherwise 0).

In the syllogisms, deriving a logically correct conclusion and agreement with a logically correct conclusion were scored 1; whereas deriving a conclusion that did not follow from the premises or agreement with a logically incorrect conclusion were scored 0.

Results and Discussion

In presenting the results of Study 2, we focus on (1) students’ understanding of logical necessity within algebraic reasoning tasks versus their algebraic transformational proficiency, (2) their performance on verbal syllogisms that require understanding of logical necessity, and (3) relations between understanding of logical necessity in algebraic and verbal deductive reasoning tasks. Because the analysis did not reveal significant gender differences, in all

<table>
<thead>
<tr>
<th>Problems and Response Categories</th>
<th>Younger Group</th>
<th>Older Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE</td>
<td>RNE</td>
</tr>
<tr>
<td><strong>Victor problem:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understood equations</td>
<td>81</td>
<td>72</td>
</tr>
<tr>
<td>Disagreed with Victor (proof is sufficient)</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>Absolutely certain in the deductively derived conclusion</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td><strong>Girl problem:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used correct equations</td>
<td>23</td>
<td>24</td>
</tr>
<tr>
<td>Requested numerical confirmations</td>
<td>84</td>
<td>89</td>
</tr>
<tr>
<td>Absolutely certain in the deductively derived conclusion</td>
<td>26</td>
<td>25</td>
</tr>
</tbody>
</table>
subsequent analyses, boys and girls were treated as one group. Table 8 presents percentages of students’ responses to questions posed in the Victor and the Girl problems, and direction of effects.

Data presented in the table clearly indicate that equivalent percentages of students in both groups exhibited algebraic proficiency—understanding and writing of correct equations. The groups differed markedly, however, in students’ certainty about deductively derived conclusions and in their beliefs about the sufficiency of proof and the importance of numerical confirmations of deductively derived conclusions. These same data aggregated across the Girl and Victor problems are graphically presented in Figure 3. The figure presents average composite scores. For each participant the composite score was a sum of absolute certainty about deductively derived conclusions in both algebraic reasoning tasks (2 if certain in both tasks, 1 if certain in one task, otherwise 0), the necessity of numerical confirmations of proved statements in the Girl problem (1 if they did not require confirmations, otherwise 0); and disagreement with Victor because the proof is sufficient (1 if disagreement was reported, otherwise 0). The data in the figure clearly point to differences in understanding of logical necessity between older RE and RNE students.

It could be counterargued, however, that the data in Figure 3 simply reflect the fact that students in the RNE group are more skeptical and, therefore, that their overall level of certainty is lower than that of the RE students. To dispel this counterargument, data on students’ certainty under different imaginary conditions were aggregated across algebraic reasoning tasks. Two aggregate scores were calculated for each group: the first score represented the proportion of students who expressed absolute certainty about the conclusion under imaginary condition 1 (i.e., the conclusion was proved), whereas the second represented the proportion of students who expressed absolute certainty about the conclusion under imaginary condition 2 (i.e., the conclusion was not proved but was instantiated more than a hundred times). These aggregate scores are presented in Figure 4.

Data presented in Figure 4 clearly indicate that the majority of older students in both groups were capable of being certain; however, their sense of certainty was derived from different sources. For the majority of older RE students, their certainty in the conclusion was justified by the deductive proof (and not by frequent instantiation), whereas for the majority of the older RNE students, their certainty was justified by a large number of supporting numerical examples (and not by the deductive proof). Note that a large number of supporting examples justified the certainty of the younger students in both groups.

The overall pattern of findings suggests that students in both groups differed in sources of justifica-

2. Aggregation of scores was carried out via calculation of percentage of students who expressed absolute certainty in the conclusion in, at least, one reasoning task.
tion for their beliefs rather than in their algebraic transformational proficiency. Older students in the RE group exhibited a markedly better understanding of the logical necessity of deductively derived algebraic conclusions than older students in the RNE group. (Note that no differences were found for the younger students.)

The results were similar for verbal deductive reasoning tasks. For the younger students, within-group average composite scores (5.1 in the RE group and 5.1 in the RNE group) did not differ significantly; whereas for the older students, these scores (9.7 in the RE group and 8.2 in the RNE group) differed significantly. $F(1, 104) = 5.32, p = .02$.

To determine relations between understanding of logical necessity in algebraic tasks and in verbal reasoning tasks, the correlation between the composite performance on the syllogisms and the composite score of understanding of logical necessity in algebraic tasks was calculated. The data indicate that (1) students’ performance on the syllogisms was related to (2) their understanding of logical necessity in algebraic reasoning tasks, $r_{xy} = .58, p < .001$.

The results of this study provide further evidence that allows us to link understanding of logical necessity to instruction. For the majority of older RE students, deductive inference was sufficient for being certain about the conclusion, whereas for the majority of older RNE students, frequent instantiation was a sufficient condition for such certainty. Therefore it can be concluded that there are many older RNE students who do not understand logical necessity, confusing proof with evidence and evidence with proof. These differences, however, were not present in the younger groups, where the majority of students expressed certainty only in frequently instantiated (but not deductively derived) conclusions. It seems particularly important that these differences were present across domains of reasoning.

It is also worth noting that there is a high correlation between understanding of logical necessity in algebraic reasoning and in verbal deductive reasoning, which suggests the possibility of a transfer of understanding of logical necessity from algebraic to verbal deductive reasoning. Indeed, because students had had no formal training in first-order logic and no differences were found for younger groups, the possibility of effects of verbal deductive reasoning on algebraic reasoning does not seem plausible. There are two other possibilities to explain the correlation: (1) the effects of algebraic reasoning on verbal deductive reasoning or (2) the presence of a third factor that systematically affects reasoning across the domains of reasoning. In our view, such a third factor could be intelligence, coursework, or the type of algebraic instruction. However, given the controls for intelligence and coursework, instruction seems to be a likely candidate for this third factor. Thus both 1 and 2 seem to lend support to the notion of transfer of understanding of logical necessity from algebraic to
verbal deductive reasoning. A future experimental study can bring direct evidence as to whether such transfer is indeed possible.

GENERAL DISCUSSION

Data from both studies demonstrate that (1) across the groups, very few younger students exhibited an understanding of logical necessity in either the domain of algebraic or verbal deductive reasoning; (2) older students in the RE group exhibited understanding of logical necessity across the domains of reasoning and exhibited a profound developmental progression in their understanding of logical necessity; (3) very few students in the English groups exhibited understanding of logical necessity and did not develop such understanding with age; (4) transformational proficiency is not sufficient for the development of understanding of logical necessity; and (5) understanding of logical necessity in algebraic reasoning was associated with understanding of logical necessity in verbal deductive reasoning. These data fit the hypothesis, further supporting the view that instruction is an important factor in the development of understanding of logical necessity and casting doubt on the view that understanding of logical necessity develops naturally.

Significant differences in understanding of logical necessity between the RE group and all other groups, and significant and systematic developmental gains in this group point to a link between type of algebra instruction and the development of the metalevel of deductive reasoning. The presented studies demonstrate that prolonged formal instruction may have long-term effects on the development of abstract deductive reasoning, amplifying the development of the metalevel of deduction across domains of reasoning. The studies also bring evidence that transformational proficiency is not sufficient for the development of understanding of logical necessity. The findings of both studies seem to support the view that instruction that fosters the acquisition of abstract concepts and the symbolism necessary for representing these concepts, amplifies the cognitive development of the child (Luria, 1976, 1982; Vygotsky, 1984; Vygotsky & Luria, 1993). Future studies, particularly a longitudinal study that introduces instruction in metacomponents of any formal system (algebra, language, or logic), controls potential confounds, and employs multiple measures of the development of metacomponents and transformational components of abstract reasoning over protracted periods of time, would help to further our understanding of the role of formal instruction in the development of abstract deductive reasoning.

The results allow us to suggest the following conclusions: (1) Instruction with a specific emphasis on metalevel abilities of algebraic deductive reasoning is related to greater developmental progression in understanding of logical necessity in algebraic and verbal deductive reasoning. (2) Many students do not develop an understanding of logical necessity naturally. (3) Understanding of logical necessity in algebraic reasoning is related to that in verbal deductive reasoning. (4) Transformational proficiency is not sufficient for the development of understanding of logical necessity.

APPENDIX

I. Categorical Syllogisms for Which Students Were Asked to Select Conclusions

The asterisk identifies the conclusion that logically follows from the premises.

Instructions: Please read the tasks carefully. Some of them assert absurd things, so, please, pay attention only to their logical form. Assume the first two sentences are true. Make a conclusion that logically follows from these sentences by choosing 1, 2, 3, or 4. Please circle your answer.

1. All fahmost numbers can be divided evenly by 8.
   26 is a fahmost number.
   Therefore:
   1. 26 must be a fahmost number.
   2. 26 is an exception to the rule.
   3. * 26 can be divided evenly by 8.
   4. It is probably true that all fahmost numbers cannot be divided evenly by 8.

2. All felismo animals are not predators.
   The lion is a felismo animal.
   Therefore:
   1. The lion must not be a felismo animal.
   2. The lion is an exception to the rule.
   3. * The lion is not a predator.
   4. It is probably not true that all felismo animals are not predators.

3. All demuted cities are not in England.
   London is a demuted city.
   Therefore:
   1. London must not be a demuted city.
   2. London is an exception to the rule.
   4. It is probably not true that all demuted cities are not in England.

4. All prosorial chemical elements are gases.
   Gold is a prosorial chemical element.
   Therefore:
   1. Gold must not be a prosorial element.
   2. Gold is an exception to the rule.
   3. * Gold is a gas.
   4. It is probably true that all prosorial chemical elements are not gases.
II. Categorical Syllogisms for Which Students Were Asked to Endorse Conclusions

Believability of conclusions measured on a 0–6 scale (from 0, totally unbelievable, to 6, totally believable) are in parentheses.

Instructions: Please read the tasks carefully. Some of them assert absurd things, so, please, pay attention only to their logical form. For all the tasks, assume the first two sentences (in bold) are true. Does the conclusion follow from the sentences? Encircle YES, if the conclusion follows, or NO, if the conclusion does not follow.

A. Believable Conclusions That Do Not Logically Follow from the Premises

1. (Believability = 5.5)
   All fractions are equipotential.
   \( \frac{a}{b} \) is equipotential.
   Therefore, \( \frac{a}{b} \) is a fraction.
   YES NO

2. (Believability = 4.3)
   All mammals have a highly organized brain.
   The elephant’s brain is highly organized.
   Therefore, the elephant is a mammal.
   YES NO

3. (Believability = 5.2)
   Some teachers are women.
   Jo Brown is a teacher.
   Therefore, Jo Brown is a woman.
   YES NO

B. Unbelievable Conclusions That Logically Follow from the Premises

4. (Believability = 1.5)
   All expositional expressions are always equal to zero.
   \( x + 1 \) is an expositional expression.
   Therefore \( x + 1 \) is always equal to zero.
   YES NO

5. (Believability = 0.4)
   All plants that have bolded skins are cucumbers.
   A strawberry has a bolded skin.
   Therefore, a strawberry is a cucumber.
   YES NO

6. (Believability = 0.9)
   All compounds that have gnoceological elements are flammable.
   Water has a gnoceological element.
   Therefore water is flammable.
   YES NO

C. Unbelievable Conclusions That Do Not Logically Follow from the Premises

7. (Believability = 0.5)
   Some gases are lighter than water.
   Wood is lighter than water.
   Therefore, wood is a gas.
   YES NO

8. (Believability = 0.6)
   Some negative numbers are real numbers.
   177 is a real number.
   Therefore, 177 is a negative number.
   YES NO

9. (Believability = 0.4)
   A wolf preys on little animals.
   An eagle preys on little animals.
   Therefore, an eagle is a wolf.
   YES NO

ACKNOWLEDGMENTS

Portions of this manuscript have been presented at the 1995 meeting of the Society for Research in Child Development and at the 1995 annual convention of the American Psychological Association. The reported studies have been supported by the Ohio State University, Johann Jacobs Foundation of Switzerland, and the University of Delaware General University Research Fund. The first study is a part of the dissertation research conducted by Anne Morris while at the Ohio State University. The authors would like to thank David Moshman, Anita Woolfolk Hoy, Michael Kamil, Michael Newman, Robert Siegler, and Gerald Winer, and the anonymous reviewer for their thoughtful comments on the earlier versions of the manuscript.

ADDRESSES AND AFFILIATIONS

Corresponding authors: Anne K. Morris, Department of Educational Development, University of Delaware, Newark, DE 19716; e-mail: abmorriss@brahms.udel.edu. Vladimir M. Sloutsky, College of Education and Center for Cognitive Science, 257 Arps Hall, 1945 North High Street, Ohio State University, Columbus, OH 43210; e-mail: sloutsky.1@osu.edu. The authors are listed alphabetically. Please address correspondence to either author.

REFERENCES


Braine, M. (1990). The “natural logic” approach to reason-


Moshman, D. (1990). The development of metalexical un-


