Generality of Academic Self-Efficacy Judgments: Evidence of Hierarchical Relations

Mimi Bong
Ewha Womans University and Dongguk University

The generality of academic self-efficacy judgments was examined among 588 high school students. Students rated their confidence for solving 42 problems in English, Spanish, U.S. history, algebra, geometry, and chemistry. Confirmatory factor analyses showed that students’ efficacy perceptions prevailed beyond the boundaries of specific problems. The 1st-order model with a separate self-efficacy factor for each school subject displayed the best fit. Verbal and Quantitative Academic Self-Efficacy illustrated the relations among the 1st-order factors better than General Academic Self-Efficacy. The generality of academic self-efficacy partly depended on the degree of perceived similarity among tasks. When asked to rate their efficacy toward 8 pairs of isomorphic algebra and physics problems, students reported more comparable strengths of self-efficacy as they perceived greater similarity between the problems.

Academic self-efficacy refers to one’s perceived capability to perform given academic tasks at desired levels (Schunk, 1991). Students with a strong sense of academic self-efficacy have been proven to willingly undertake challenging tasks (Bandura & Schunk, 1981), expend greater effort for accomplishing a given task (Salomon, 1984; Schunk, 1983), persist longer in the presence of difficulties (Bandura & Schunk, 1981; Schunk, 1982), demonstrate lower levels of anxiety (Meece, Wigfield, & Eccles, 1990; Pintrich & De Groot, 1990), use more effective learning strategies (Pintrich & De Groot, 1990), and self-regulate better than others (Zimmerman, Bandura, & Martinez-Pons, 1992; Zimmerman & Martinez-Pons, 1990). Academic self-efficacy is often conceptualized as a domain-specific construct, and its relationships with various achievement indexes have frequently been probed in the context of carrying out a specific task of interest (e.g., subtraction; Bandura & Schunk, 1981). Recently, however, several researchers began to suspect that academic self-efficacy possesses a certain degree of generality across activities or domains (Pajares, 1996b; Schunk, 1991).

In fact, Bandura (1986) suggested that self-efficacy judgments can vary along the dimensions of level, generality, and strength. People’s efficacy perception may involve only simple and easy tasks or include even the most demanding ones within a particular domain (level); individuals may judge themselves to be effective only in a restricted range of domains or consider themselves efficacious across a wide range of activities and situations (generality); persons with a strong sense of personal competence remain steadfast on occasional failures, whereas those with weak percepts of efficacy are quick to resign with self-debilitating expectations (strength).

To date, most academic self-efficacy studies have used measures that focus on the strength of efficacy beliefs. Although this is a valid practice, it is noteworthy that only few studies have systematically investigated the generality of academic self-efficacy perceptions (e.g., Zimmerman & Ringle, 1981). When we consider that self-judged efficacy has been proven repeatedly to wield significant impact on so many facets of students’ academic quest, it is certainly of interest to find out whether and how much students with firm (or weak) percepts of efficacy in one domain face other domains of academic functioning with equal certitude. For example, would Schunk’s (1982, 1983) students now armed with stronger self-efficacy toward subtraction confront division problems with increased self-assurance? Or, better yet, do they feel more confident in solving mathematics problems as a whole?

Self-efficacy was originally conceptualized as a predictive measure of behavioral change (Bandura, 1977, 1986; Zimmerman, 1995). Therefore, it is most useful when the specificity of efficacy measurement corresponds to the specificity of target performance (Pajares, 1996a; Pajares & Miller, 1995). To this end, we should be able to demonstrate first that academic self-efficacy indeed displays some gener-
ality over the confines of any single task. The primary purpose of the present investigation is to examine the degree of generality of academic self-efficacy judgments and explore the relations among domain-specific perceptions of academic self-efficacy.

The Role of Perceived Similarity in Self-Efficacy Generalization

Bandura (1986) stated that

Once established, enhanced self-efficacy tends to generalize to other situations . . . As a result, behavioral functioning may improve across a wide range of activities. However, the generalization effects occur most predictably in activities that are most similar to those in which self-efficacy was enhanced. (p. 399)

There is evidence that individuals encounter multiple tasks with comparable self-confidence when those tasks are perceived to be similar. In one experiment reported by Bandura, Adams, and Beyer (1977), adults suffering from chronic snake phobia received remedial treatment. Following successful treatment using a boa constrictor, participants rated their perceived efficacy for executing other tasks (e.g., coping with snakes in natural settings, coping with other feared animals, handling difficult social situations) that varied along a dimension of similarity to the original threat (i.e., coping with snakes in experimental settings). They displayed significantly enhanced self-efficacy for all the generalization tasks. Changes in efficacy, however, decreased in magnitude as purported similarity of the generalization tasks to the original threat progressively declined (see t values reported for within-group comparison in Bandura et al., 1977, Table 1, p. 130).

Shell, Murphy, and Brumming's (1989) study also hints to the proposed relationship between perceived similarity and the generality of self-efficacy. College students reported two types of efficacy in both reading and writing: component skill efficacy, which referred to confidence in performing each of the skills listed, and task efficacy, which referred to confidence in being able to read and understand (or to write for successfully communicating) each of the tasks listed. A canonical correlation analysis resulted in a single dimension underlying reading and writing such that skilled readers also tended to be proficient writers and that they exhibited a strong sense of efficacy in both areas. More interesting, the correlations between reading and writing component skill efficacy (r = .85) and between reading and writing task efficacy (r = .75) were greater than those between component skill and task efficacy in the same area (r = .65 in reading; r = .62 in writing). An examination of the efficacy rating items reveals that each of the component skill and task efficacy items in reading (e.g., read a letter from a friend or from a family member, recognize grammatically correct sentence structure) bears remarkable resemblance to the corresponding item in writing (e.g., write a letter to a friend or to a family member, write a simple sentence with proper punctuation and grammatical structure).

In school, students are likely to consider a set of academic tasks to be similar or different. If one has already developed a strong sense of efficacy toward a particular task, he or she would confront similar tasks with reasonable strengths of personal competence based on his or her prior mastery experiences with the original task (Bandura, 1986; Bandura et al., 1977). By the same token, one's belief that the task at hand is analogous to the one with which he or she has been unsuccessful in the past could severely hamper self-judged efficacy for the given task. Results from the present investigation are expected to describe the existing patterns of generality in academic self-efficacy judgments and provide researchers with empirical justification for assessing students' efficacy beliefs at differing levels of specificity depending on their predictive and explanatory goals.

Specifically, the following research questions were explored: (a) Do high school students hold generalized perceptions of their academic capability beyond the boundaries of specific tasks, school subjects, or both? (b) If self-judged efficacy demonstrates some generality over the confines of particular school subjects, is the generality greater among verbal subjects or math-related subjects? (c) Is the generality of academic self-efficacy judgments great enough to warrant assessment of "general" academic self-efficacy that encompasses several discrete domains? and (d) How well can perceived task similarity predict academic self-efficacy judgment?

Method

Participants

Five hundred and eighty-eight students (49% male, 51% female) from four high schools in Los Angeles County, California, participated. Ages ranged from 15 years 7 months to 21 years (M = 17.9 years). A majority of the students were in Grades 11 (23%) and 12 (76%). Only four students were from Grade 10 (1%). Ethnic composition was as follows: 20% White, 7% African American, 55% Hispanic, 16% Asian, and 1% Native American. Data were collected from students enrolled in the following classes: 1 advanced-placement (AP) physics, 1 AP economics, 22 regular economics and American government, and 5 regular chemistry classes. On average, participants had taken 5.5 English (Mdn = 6.0), 3.3 Spanish (Mdn = 4.0), 4.1 social studies (Mdn = 4.0), 2.9 algebra (Mdn = 2.0), 1.6 geometry (Mdn = 2.0), and 1.2 chemistry (Mdn = 2.0) courses at the time of the survey.

Measures and Procedures

Academic self-efficacy judgments. Six school subjects were selected for assessing students' academic self-efficacy perception: English, Spanish, U.S. history, algebra, geometry, and chemistry. Seven representative problems for each school subject were prepared from Scholastic Aptitude Test (SAT) preparatory booklets (Brownstein, Weiner, & Green, 1994; College Entrance Examination Board and Educational Testing Service, 1994; see Appendix A for sample problems). There were three main reasons for presenting SAT problems: (a) Problems needed to be independent of any particular text or instructional method because the sample was being drawn from four different high schools in two school districts; (b) a vast majority of the participants were in Grades 11 and 12 who, teachers indicated, were familiar with types and procedures of SAT problems; and (c) presenting specific problems that could function as concrete anchor points for judging academic
self-efficacy was deemed more appropriate rather than presenting merely verbal descriptions of various tasks that tended to be more general, given the purpose of the present study. A problem was considered for inclusion only when the reported percentage of students answering it correctly was between 50 and 80 whenever such information was available.

A final set contained 42 problems from the six academic subjects. The problems were randomly interspersed with each other to prevent a potential “set effect” that might influence students’ perception of problem similarity. It was suspected that if problems from the same school subject were to be presented consecutively, students could perceive greater similarity among them, which, in turn, might result in higher intercorrelations among self-efficacy ratings. The main focus of the current investigation is to find out the existing patterns of generality in students’ academic self-efficacy beliefs without any manipulation. Therefore, it would be inappropriate to influence students’ similarity perception in any way if such practice alters what would have been obtained otherwise. In addition, Marsh, Walker, and Debus (1991) have warned that the common practice of efficacy researchers to present items together by each school subject could be responsible for the observed differences between academic self-efficacy and self-concept. Accordingly, problems were arranged in a random order regardless of the particular school subjects to which they belonged.

Each problem was presented by an overhead projector for approximately 10 to 20 s. The duration of exposure had been set in advance for each problem so that it would be long enough to recognize the type of a given problem but too short to actually solve it. Students judged privately their self-perceived capability for solving correctly each type of the problems on a scale ranging from 0 to 100 in 10-unit intervals. The efficacy rating scale had the following verbal descriptors to help students understand more clearly what each number represented: 0 (not sure), 40 (maybe), 70 (pretty sure), and 100 (real sure). This type of self-efficacy scale has been commonly used in previous research (e.g., Bandura & Schunk, 1981; Schunk, 1982, 1983).

Perceived similarity and the generality of academic self-efficacy judgments. To test the hypothesized relationship between similarity perception and the generality of academic self-efficacy, I asked students to report their confidence for solving successfully 16 additional problems adapted from 8 original problems used in Bassok and Holyoak (1989, Experiment 1). Half of the problems involved arithmetic progression taught in algebra and the other half involved constant acceleration taught in physics. The two types of problems are isomorphic such that knowledge of one principle (e.g., arithmetic progression) should enable one to solve problems involving the other principle (e.g., constant acceleration). In Gick and Holyoak’s term (1987), they are structurally similar. In Bassok and Holyoak’s experiment, more than half of the students who had been taught with arithmetic progression problems were able to perceive the structural parallelism in constant acceleration problems and apply what they had learned in algebra to solving physics problems. However, those who had been trained with constant acceleration principle in physics failed to perceive the systematic correspondence or to transfer their physics knowledge to novel algebra problems.

This finding demonstrates that students are not particularly adept at perceiving the structural similarity between the two classes of problems. Hence, if perception of similarity among tasks actually plays a significant role in determining the generality of academic self-efficacy judgments, a dramatic difference will be obtained between students who successfully discern the parallelism and those who do not. Specifically, it was expected that students would face the problems with similar strengths of assurance based on the same prior experiences if they were tuned to the similarity between them; if not, students would estimate their personal competence for each problem separately on the basis of different mastery experiences.

In the present study, eight problem pairs were constructed for comparison (see Appendix B for sample problem pairs). The pairs consisted of arithmetic progression problems, constant acceleration problems, or both. Each of the two problems (Problems A and B) in a given pair was first presented separately for approximately 20 to 40 s. For arithmetic–physics problem pairs, the order of presentation was counterbalanced so that half of the pairs presented arithmetic problems first, whereas the other half presented physics problems first. Students rated their self-efficacy for successfully solving each type of the problems on a scale ranging from 0 (not sure) to 100 (real sure). Both problems in the given pair were then presented together for another 20 s. At this time, students were asked to judge perceived similarity between the two problems on a scale ranging from 1 (not similar at all) to 7 (very similar).

Results

Means and standard deviations of academic self-efficacy scales for the six school subjects as a function of grade level, gender, and ethnicity are presented in Table 1. Different superscripts denote significant difference between group means. A 2 (grade level) X 2 (gender) X 4 (ethnicity) analysis of variance (ANOVA) detected no significant 2- or 3-way interaction. All academic self-efficacy scales for the six school subjects demonstrated acceptable levels of reliability. The standardized Cronbach’s alphas ranged from .86 to .97 (Mdn = .91; see Table 2).

Confirmatory Factor Analyses

Tests of first-order factor models with different degrees of generality. Six first-order confirmatory factor analysis (CFA) models were specified to test different degrees of generality of academic self-efficacy judgments. Eighteen measured variables were created by averaging students’ self-efficacy ratings for two to three problems in each school subject. When students failed to respond to any of the problems, average scores based on available responses were computed. Ten students failed to respond to all of the problems comprising one or more measured variables and were excluded from further analyses. Table 3 reports intercorrelations among the eighteen measured variables.

All CFAs were performed with the EQS (Bentler, 1992) program. The following goodness-of-fit indexes were used in determining the model fit: ratio of chi-square value to its degrees of freedom, Bentler-Bonett nonnormed fit index (NNFI), comparative fit index (CFI), and average absolute standardized residuals. The ratio of chi-square value to its degrees of freedom was preferred over the exact probability level because the chi-square test is known to be biased against a large sample size. The NNFI and CFI can be interpreted roughly as indicating the amount of variance in the observed data explained by the particular model. Values of .90 or greater for either index are commonly taken as evidence of adequate model fit. The NNFI is a mathematical equivalent of the Tucker-Lewis fit index reported by the LISREL program (Jöreskog & Sörbom, 1989). Average absolute standardized residuals is an index that represents
the magnitude of deviation in the model's capability to reproduce the empirical data.

Table 4 presents the model descriptions and goodness-of-fit indexes. Model 1 hypothesized that students' perceptions of their academic capability were analogous within a set of school subjects that primarily require either verbal skills in English, verbal skills in Spanish, or quantitative skills. Model 1 did not fit the data well. Models 2 and 3 presumed greater specificity of academic self-efficacy judgments compared with Model 1. Model 2 hypothesized a reduced degree of generality in math- and science-related areas. Model 3 also tested a four-factor structure by specifying individual self-efficacy factors for each of the verbal subjects. The goodness-of-fit indexes for both models were better than those for Model 1 but still failed to delineate the empirical data adequately (see Table 4).

Next, two models hypothesized five first-order academic self-efficacy factors. Model 4 was based on the presumption that academic self-efficacy judgments were being generalized between English and U.S. history because both subjects demand good verbal skills in English. Model 4 assumed less generality in the quantitative area, specifying a separate factor for each of the math and science subjects. In contrast, Model 5 postulated English, Spanish, U.S. History, Math, and Chemistry Academic Self-Efficacy under the premise that high school students possess generalized perception of academic self-efficacy within math subjects—algebra and geometry. Both models fit the data reasonably well with Model 5 demonstrating a better fit than Model 4 (see Table 4).

The last first-order CFA model (Model 6) posited a separate first-order factor for each of the six school subjects. The fit of this model was most satisfactory with all the fit indexes exceeding those of other models (see Table 4). All the factor loadings and error variances for Model 6 were statistically significant ($p < .05$). The only difference in factor structure between Models 5 and 6 was the integration of (Model 5), or the separation between (Model 6), the Algebra and Geometry Academic Self-Efficacy factors. Values for the standardized factor loadings ranged between .680 and .976 for Model 5 ($M = .860$, $Mdn = .863$) and between .678 and .976 for Model 6 ($M = .867$, $Mdn = .875$). These values suggest that each of the first-order academic self-efficacy factors was clearly defined in both models. The magnitude of factor loadings would not have been as substantial as observed if students' assessment of their own competence varied a great deal across different types of problems presented in each school subject. Therefore, students appeared to hold more or less generalized percep-

### Table 1

<table>
<thead>
<tr>
<th>Scale</th>
<th>Grade level</th>
<th>Gender</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 11</td>
<td>Grade 12</td>
<td>Boy</td>
</tr>
<tr>
<td>English</td>
<td>$M=75.8_4$</td>
<td>$M=70.9_6$</td>
<td>71.1</td>
</tr>
<tr>
<td>$SD=16.1$</td>
<td>$SD=18.6$</td>
<td>$SD=19.0$</td>
<td>17.3</td>
</tr>
<tr>
<td>Spanish</td>
<td>$M=71.6$</td>
<td>$M=66.6$</td>
<td>65.0$ _a$</td>
</tr>
<tr>
<td>$SD=29.2$</td>
<td>$SD=33.4$</td>
<td>$SD=31.7$</td>
<td>33.0</td>
</tr>
<tr>
<td>U.S. history</td>
<td>$M=70.9_6$</td>
<td>$M=65.3_6$</td>
<td>69.8$ _a$</td>
</tr>
<tr>
<td>$SD=20.7$</td>
<td>$SD=20.2$</td>
<td>$SD=19.5$</td>
<td>21.0</td>
</tr>
<tr>
<td>Algebra</td>
<td>$M=63.1$</td>
<td>$M=58.8$</td>
<td>60.7</td>
</tr>
<tr>
<td>$SD=26.0$</td>
<td>$SD=25.0$</td>
<td>$SD=26.2$</td>
<td>24.5</td>
</tr>
<tr>
<td>Geometry</td>
<td>$M=61.9_4$</td>
<td>$M=56.6_4$</td>
<td>60.0$ _a$</td>
</tr>
<tr>
<td>$SD=23.9$</td>
<td>$SD=25.8$</td>
<td>$SD=26.6$</td>
<td>24.0</td>
</tr>
<tr>
<td>Chemistry</td>
<td>$M=57.5_4$</td>
<td>$M=48.2_4$</td>
<td>51.6</td>
</tr>
<tr>
<td>$SD=24.9$</td>
<td>$SD=22.7$</td>
<td>$SD=24.9$</td>
<td>22.0</td>
</tr>
</tbody>
</table>

*Note.* Means with different subscripts differ significantly at $p < .05$ within each category. In the case of ethnicity, Scheffé multiple comparison tests were conducted after significant main effects were detected.

The means with different subscripts differ significantly at $p < .05$ within each category. In the case of ethnicity, Scheffé multiple comparison tests were conducted after significant main effects were detected.

### Table 2

<table>
<thead>
<tr>
<th>Scale</th>
<th>$M$</th>
<th>$SD$</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>72.07</td>
<td>18.20</td>
<td>.86</td>
</tr>
<tr>
<td>Spanish</td>
<td>67.58</td>
<td>22.52</td>
<td>.97</td>
</tr>
<tr>
<td>U.S. history</td>
<td>66.44</td>
<td>20.47</td>
<td>.90</td>
</tr>
<tr>
<td>Algebra</td>
<td>59.76</td>
<td>25.35</td>
<td>.90</td>
</tr>
<tr>
<td>Geometry</td>
<td>57.83</td>
<td>25.41</td>
<td>.93</td>
</tr>
<tr>
<td>Chemistry</td>
<td>50.36</td>
<td>23.51</td>
<td>.91</td>
</tr>
</tbody>
</table>
tions of their academic capability beyond the boundary of a specific problem.

The correlations among the factors in Model 5 were modest in absolute values ($M = .41, Mdn = .48$; see Table 5). Those in Model 6 ranged from .11 to .92 ($M = .45, Mdn = .47$; see Table 6). A correlation coefficient greater than .90 can raise a question of discriminant validity. In addition, the rule of parsimony dictates that a simpler model should be taken when there is no significant difference in fit between the simpler and more complex models. Model 6 displayed a significant reduction in chi-square value over Model 5, $\Delta \chi^2(5, N = 578) = 121.909, p < .001$, furnishing the ground for six separate factors.

Hierarchical CFA. Four second-order factor structures based on Model 5 and Model 6 were fitted to the empirical data to further test the generality of academic self-efficacy judgments beyond the boundaries of specific school subjects. For example, Model 5 strongly suggests that students may hold generalized perceptions of academic self-efficacy over a subject boundary within the area of math. Although Model 6 may lack discriminant validity regarding some of its factors, it was nevertheless used as a basis for the second-order factor models because, first, a minimum of three first-order factors are needed to define a single second-order factor and second, it showed a significantly better fit to the data compared with Model 5.

Lower order factors should be sufficiently correlated among themselves to warrant testing any hierarchical model. The fit of a hierarchical model can only be as good as that of the lower order model on which it is grounded. When the fit of a hierarchical model is not significantly different from that of its lower order counterpart, the hierarchical model is preferred because it is considered more parsimonious between the two. Vispoel (1995) suggested comparing the fit of lower order models with either a correlated or uncorrelated factor structure for determining the strength and effectiveness of potential hierarchies in explaining the covariances among lower order factors. Table 4 presents the goodness-of-fit indexes for Models 5 and 6 with either a correlated or uncorrelated factor structure for determining the strength and effectiveness of potential hierarchies among lower order factors. The model fit was significantly improved by specifying correlations among the factors in both Model 5, $\Delta \chi^2(10, N = 578) = 881.334, p < .001$, and Model 6, $\Delta \chi^2(15, N = 578) = 1601.282, p < .001$.

Model 7 specified a single second-order factor, General Academic Self-Efficacy, defined by five first-order factors. Model 7 fit the data well (see Table 4). Evaluating the fit of a hierarchical model, however, necessitates an inspection of additional indicators. The NNFI and CFI can only demonstrate the ability of an entire model—including both lower and higher order factors—to explain the covariances observed among measured variables. Such indexes can be misleading in hierarchical models where lower order factors are relatively uncorrelated, rendering most of their variance unexplainable by higher order factors. Marsh and Hocevar (1985) thus recommended the use of a target coefficient (TC) together with other goodness-of-fit indexes. The TC reflects the proportion of variance in lower order factors that is explained by higher order factors, indicating exclusively the fit of hierarchies imposed.
Table 4
Descriptions of Confirmatory Factor Analysis Models Tested and Their Goodness-of-Fit Indexes

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$\chi^2$/df</th>
<th>NNFI</th>
<th>CFI</th>
<th>Res.</th>
<th>TC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>18 uncorrelated first-order factors</td>
<td>9488.065</td>
<td>153</td>
<td>62.014</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Four correlated first-order factors: Verbal-English, Verbal-Spanish, Math, and Chemistry Academic Self-Efficacy</td>
<td>1062.882</td>
<td>129</td>
<td>8.239</td>
<td>.881</td>
<td>.900</td>
<td>.033</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Five correlated first-order factors: English, Spanish, U.S. History, Math, and Chemistry Academic Self-Efficacy</td>
<td>731.905</td>
<td>125</td>
<td>5.855</td>
<td>.920</td>
<td>.935</td>
<td>.026</td>
<td>1.00</td>
</tr>
<tr>
<td>5A</td>
<td>Same as 5 except that all factors are uncorrelated</td>
<td>1613.239</td>
<td>135</td>
<td>11.950</td>
<td>.821</td>
<td>.842</td>
<td>.229</td>
<td>.000</td>
</tr>
<tr>
<td>6</td>
<td>Six correlated first-order factors: English, Spanish, U.S. History, Algebra, Geometry, and Chemistry Academic Self-Efficacy</td>
<td>609.996</td>
<td>120</td>
<td>5.083</td>
<td>.933</td>
<td>.948</td>
<td>.024</td>
<td>1.00</td>
</tr>
<tr>
<td>6A</td>
<td>Same as 6 except that all factors are uncorrelated</td>
<td>2211.278</td>
<td>135</td>
<td>16.380</td>
<td>.748</td>
<td>.778</td>
<td>.264</td>
<td>.000</td>
</tr>
<tr>
<td>8</td>
<td>One second-order factor defined by six first-order factors (English, Spanish, U.S. History, Algebra, Geometry, and Chemistry Academic Self-Efficacy)</td>
<td>873.223</td>
<td>129</td>
<td>6.769</td>
<td>.905</td>
<td>.920</td>
<td>.054</td>
<td>.836</td>
</tr>
<tr>
<td>9</td>
<td>Two correlated second-order factors: Verbal and Quantitative Academic Self-Efficacy, each defined by three first-order factors</td>
<td>715.651</td>
<td>128</td>
<td>5.591</td>
<td>.925</td>
<td>.937</td>
<td>.039</td>
<td>.934</td>
</tr>
<tr>
<td>10</td>
<td>Two correlated second-order factors: Verbal and Quantitative Academic Self-Efficacy, with first-order Chemistry Academic Self-Efficacy factor loading on both second-order factors</td>
<td>637.352</td>
<td>127</td>
<td>5.019</td>
<td>.934</td>
<td>.945</td>
<td>.028</td>
<td>.983</td>
</tr>
</tbody>
</table>

Note. N = 578. NNFI = Bentler-Bonnett nonnormed fit index; CFI = comparative fit index; Res. = average absolute standardized residuals; TC = target coefficient.

The TC for Model 7 was .895, indicating that more than 89% of the covariances among the five first-order factors was accounted for by the General Academic Self-Efficacy factor. An inspection of the disturbance terms revealed, however, that the variance of the first-order Spanish Academic Self-Efficacy factor remained virtually unexplained by the second-order factor (see Figure 1). Whereas General Academic Self-Efficacy accounted for over 50% of the variance in the first-order English, U.S. History, Math, and Chemistry Academic Self-Efficacy factors, it explained less than 5% of the variance in Spanish Academic Self-Efficacy.

Table 5
Correlations Among Five First-Order Academic Self-Efficacy Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. English Academic Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Spanish Academic Self-Efficacy</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Math Academic Self-Efficacy</td>
<td>.50</td>
<td>.11</td>
<td>.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Chemistry Academic Self-Efficacy</td>
<td>.54</td>
<td>.20</td>
<td>.63</td>
<td>.71</td>
<td></td>
</tr>
</tbody>
</table>

Note. All ps < .05.

Table 6
Correlations Among Six First-Order Academic Self-Efficacy Factors

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. English Academic Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Spanish Academic Self-Efficacy</td>
<td></td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Algebra Academic Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td>.52</td>
<td>.11</td>
<td>.44</td>
</tr>
<tr>
<td>5. Geometry Academic Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.47</td>
<td>.10</td>
</tr>
<tr>
<td>6. Chemistry Academic Self-Efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.54</td>
</tr>
</tbody>
</table>

Note. All ps < .05.
Figure 1. Model 7 with one second-order Academic Self-Efficacy factor defined by five first-order factors. All paths are significant at $p < .05$.

This can be taken as evidence that more than a single general second-order factor is needed to adequately represent the data.

Next, two second-order CFA models based on six first-order factors were explored and compared. Model 8 posited a single second-order General Academic Self-Efficacy factor as in Model 7, whereas Model 9 specified two correlated second-order factors, Verbal and Quantitative Academic Self-Efficacy. Model 9 fit the data noticeably better than Model 8 (see Table 4). In particular, Model 9 was capable of accounting for about 93% of the covariances among the six first-order factors (TC = .934), which was almost 10% more than what Model 8 could represent (TC = .836). All the goodness-of-fit indexes of Model 9 also exceeded those of Model 7.

Figures 2 and 3 depict the factor structures of Models 8 and 9, respectively. As can be seen, Model 9 brought a sizable reduction in the disturbance terms of English (.679 to .274) and U.S. History Academic Self-Efficacy (.717 to .339) from Model 8, demonstrating the necessity of a separate Verbal second-order factor apart from Quantitative Academic Self-Efficacy. The two second-order factors were significantly and positively correlated with each other (.625). Still, over 96% of the variance in the first-order Spanish Academic Self-Efficacy factor remained unexplained. It is conjectured that at least three second-order factors, Verbal-English, Verbal-Spanish, and Quantitative Academic Self-Efficacy, are needed to adequately represent the covariances among the six first-order factors examined in the present study.

The Lagrange multiplier tests conducted on Model 9 further suggested that a path be opened between the first-order Chemistry Academic Self-Efficacy factor and the second-order Verbal Academic Self-Efficacy factor. Tables 5 and 6 indeed show that Chemistry Academic Self-Efficacy demonstrated correlation coefficients of .54 and .63 with English and U.S. History Academic Self-Efficacy, respectively. When the additional parameter was incorporated in Model 10, a significant drop in chi-square value from Model 9 was observed, $\Delta \chi^2(1, N = 578) = 78.299, p < .001$. All the goodness-of-fit indexes also attested to the relative superiority of Model 10 over Models 7, 8, and 9 (see Table 4). The most striking evidence, perhaps, was found in the TC, which represented that over 98% of the covariances among the six first-order factors was accounted for by the hierarchy prescribed by Model 10. When we recall that the fit of a hierarchical model cannot surpass that of its lower order counterpart, this leaves little room for further improvement. Figure 4 illustrates that Chemistry Academic Self-Efficacy is equally related to the second-order Verbal (.437) and Quantitative (.474) Academic Self-Efficacy factors.

Degrees of generality in verbal and quantitative domains. An examination of factor correlations in Model 5 (see Table 5) and Model 6 (see Table 6) discloses higher intercorrelations of self-efficacy perceptions among math and science subjects than verbal subjects. The unexpectedly low correlations of Spanish Academic Self-Efficacy with other verbal self-efficacy factors are the major reason for the reduced generality of academic self-efficacy within the verbal area. In Figure 3, it can be seen that the second-order Quantitative
Academic Self-Efficacy displays a significant and tangible relationship with each of the first-order Algebra, Geometry, and Chemistry Academic Self-Efficacy factors. In other words, the three first-order academic self-efficacy factors are significantly correlated among themselves, sharing a large amount of their variance through Quantitative Academic Self-Efficacy. Likewise, the first-order English and U.S. History Academic Self-Efficacy factors are linked to each other through the second-order Verbal Academic Self-Efficacy factor. Although Verbal Academic Self-Efficacy is also significantly related to Spanish Academic Self-Efficacy, the magnitude of their relationship is rather negligible compared with the other paths. Figure 4 shows essentially the same pattern of relationships with the exception of Chemistry Academic Self-Efficacy loading on both of the second-order factors. Overall, it demonstrates a lack of generality of academic self-efficacy judgments between Spanish and other English-based verbal subjects.

Relationships Between Similarity Ratings and Academic Self-Efficacy Judgments

In light of the above results supporting the generality of academic self-efficacy judgments beyond a specific task and school subject, it is of great interest to find out factors that potentially wield influence on such generality. The present investigation hypothesized that perceived similarity among tasks would be positively related to the generality of academic self-efficacy perceptions. Analyses of the algebra and physics problem pairs adapted from Bassok and Holyoak (1989) provided some evidence for the proposed relationship. It will be recalled that each of the problem pairs consisted of algebra and physics problems that required similar solution methods. If the generality of academic self-efficacy judgments indeed depends on the perceived similarity among tasks, students should report more comparable strengths of self-efficacy for the set of problems they perceive as similar.

A correlational analysis was first carried out with correlation coefficients between Problems A and B in each pair as an outcome variable and average similarity ratings as a predictor variable. The relationship was found to be in the hypothesized direction, although it failed to reach statistical significance, \( r(8) = 0.43, p > .05 \). As students' ratings of similarity between Problems A and B in the pairs increased, the correlation between their self-efficacy ratings for the two problems also increased. It is clear that the exact probability level associated with the reported correlation coefficient has been penalized by the small sample size \((N = 8; \text{eight problem pairs})\). Using a different dependent measure can overcome this problem. Specifically, self-efficacy difference scores can be computed by taking the absolute difference between students' efficacy ratings for Problems A and B in each pair. Greater generality of academic self-efficacy, therefore, is represented by lower self-efficacy difference scores. Self-efficacy difference scores thus convey essentially the same information as the correlation coefficients without being constrained by the small number of problem pairs. Because difference is computed between two independent ratings of self-efficacy, scores so obtained do not necessarily suffer from unreliability. Across the eight problem pairs, self-efficacy difference scores demonstrated a negative relationship with perceived similarity ratings, \( r(575) = -0.17, p < .001 \). Mean difference between self-efficacy ratings for Problems A and B in the eight pairs decreased as average similarity ratings between them increased. Still, this method requires aggregating both predictor and outcome variables at the individual student level.

Hierarchical linear modeling (HLM) permits a more proper analysis of the nested data (Bryk & Raudenbush, 1992; Bryk, Raudenbush, & Congdon, 1996). In the present study, each of the 588 students furnished similarity ratings and self-efficacy difference scores for each of the eight problem pairs. Hence, there are 4,704 potential data points (588 students \( \times \) 8 problem pairs) from which to estimate the relation between perceived similarity and self-efficacy difference scores. However, eight ratings of each measure were provided by the same individual, posing a threat to the independence-of-observation assumption. HLM effectively deals with such nested data structures without aggregating any of the measures. Specifying a two-level model also affords partitioning of the variance based on problem pairs as well as individual students.

A base-model with no Level 1 or Level 2 predictor was first specified with individual students as Level 2 units \((N = 417)\). Students with missing data on either similarity ratings or self-efficacy difference scores were excluded. Table 7 shows that about 84% of the total variance \((100.24 / [100.24 + 19.01])\) was attributable to within-student variability (i.e., problem effect). Between-student variance also comprised a significant portion of the total variance \((19.01 / [100.24 + 19.01]) = 16\%\). A random-coefficient model with similarity ratings as a Level 1 predictor disclosed that, on average, there was a significant negative relationship \((-2.03\) between perceived similarity and self-efficacy difference scores (see Table 8). As similarity ratings increased, self-efficacy difference scores between Problems A and B
decreased. Adding perceived similarity in the model also substantially reduced the problem-level variability. Roughly 17% ([100.24 - 83.34]/100.24) of the within-student variance was now accounted for by their similarity ratings. Table 8 also shows that highly significant between-student variability remains to be explained and that the relation between similarity ratings and self-efficacy difference scores differs significantly across individual students.

A two-level model was next fitted to the data with similarity ratings and class type as Level 1 and Level 2 predictor, respectively. A score of 2 was assigned to students who came from regular classes of diverse subjects, whereas a score of 1 was assigned to those who belonged to AP classes. The coefficient associated with class type was significant and positive (2.44) in the model for student means (see Table 9). It signifies that students who participated from regular classes exhibited higher self-efficacy difference scores than those from AP classes. However, the similarity-self-efficacy difference slopes did not vary significantly between the two class types (p > .05). The residual variance at Level 2 was reduced only by 2% ([21.17 - 20.76]/21.17), and there still exists significant variability to be explained. More student-related variables will be needed to fully account for the remaining variance.

**Discussion**

The role of self-efficacy beliefs in various phases of academic functioning has often been investigated within the context of a single specific task (e.g., subtraction). However, when the goal of a researcher is to predict or explain students' scholastic performance at a more general level (e.g., mathematics), he or she is obliged to assess self-efficacy that stays at the same level of generality (Bandura, 1977, 1986; Pajares, 1996a; Pajares & Miller, 1995). Validity of such practice is contingent on empirical evidence pointing to the generality of self-judged efficacy across different academic activities. Together, results from the present investigation demonstrate that academic self-efficacy judgments prevail beyond the boundaries of specific tasks and also school subjects, albeit to a lesser degree.

**A Hierarchical Structure of Academic Self-Efficacy Judgments**

The magnitude and significance of factor loadings that defined each subject-specific self-efficacy factor provide evidence that academic self-efficacy judgments transcend the confines of a single task. The present study used measured variables that were composite scores of students' self-efficacy ratings for two to three problems in each school subject. The three measured variables so created for each subject-specific factor were found to share substantial amount of the variance through their respective factor in all first-order CFA models that displayed an acceptable fit. Such results are not to be expected if students' judgment of their scholastic capability is specific enough that it varies considerably across different types of problems even among those that belong to the same school subject. It thus appears that students' perceptions of their self-efficacy toward divergent tasks are fairly consistent within the boundary of each school subject.

**Table 7**

*Hierarchical Linear Modeling Results: A Base Model*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t ratio</th>
<th>Variance component</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed: Average student mean</td>
<td>7.86</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random: Student mean</td>
<td>19.01</td>
<td></td>
<td></td>
<td></td>
<td>416</td>
<td>1046.99</td>
<td>.000</td>
</tr>
<tr>
<td>Problem effect</td>
<td>100.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 417. Two students were excluded additionally from the base model because they did not have sufficient data for estimation.*

**Table 8**

*Effects of Perceived Similarity on Self-Efficacy Difference Scores*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t ratio</th>
<th>Variance component</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average student mean</td>
<td>7.86</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean similarity self-efficacy</td>
<td>-2.03</td>
<td>.14</td>
<td>-14.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>difference slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>21.17</td>
<td></td>
<td></td>
<td></td>
<td>414</td>
<td>1247.42</td>
<td>.000</td>
</tr>
<tr>
<td>Similarity self-efficacy</td>
<td>2.51</td>
<td></td>
<td></td>
<td></td>
<td>414</td>
<td>581.46</td>
<td>.000</td>
</tr>
<tr>
<td>difference slope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem effect</td>
<td>83.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 415. Two students were excluded additionally from the base model because they did not have sufficient data for estimation.*
Table 9
Effects of Class Type on the Relation Between Perceived Similarity and Self-Efficacy Difference Scores

<table>
<thead>
<tr>
<th>Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t ratio</th>
<th>Variance component</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model for student means</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>3.21</td>
<td>1.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class type</td>
<td>2.44</td>
<td>.96</td>
<td>2.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model for similarity self-efficacy difference slopes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-.84</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class type</td>
<td>-.62</td>
<td>.48</td>
<td>-1.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student mean</td>
<td>20.76</td>
<td>413</td>
<td>1227.45</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similarity self-efficacy difference slope</td>
<td>2.49</td>
<td>413</td>
<td>576.61</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem effect</td>
<td>83.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. N = 415. Two students from the base model were excluded because they did not have sufficient data for estimation.

Two first-order factor models, each specifying five (i.e., English, Spanish, U.S. History, Math, and Chemistry Academic Self-Efficacy) and six academic self-efficacy factors (i.e., English, Spanish, U.S. History, Algebra, Geometry, and Chemistry Academic Self-Efficacy) exhibited an especially satisfactory fit to the observed data. The data were best explained by the latter, which posited a separate self-efficacy factor for each of the six school subjects. Moreover, the six subject-specific academic self-efficacy were moderately correlated with each other, indicating that tests of greater generality were in order.

A single second-order factor, General Academic Self-Efficacy, was not able to represent adequately the observed pattern of intercorrelations among some of its first-order factors. The General Academic Self-Efficacy factor left unexplained an ample amount of the variance in English, Spanish, and U.S. History Academic Self-Efficacy. This testifies to the need for a separate second-order factor that exclusively taps into the verbal efficacy factors. Two second-order factors, Verbal and Quantitative Academic Self-Efficacy, were able to account for significantly more variance in the first-order subject-specific self-efficacy factors compared with General Academic Self-Efficacy. Even so, most of the variance in Spanish Academic Self-Efficacy still remained intact. The major reason for the greater generality of academic self-efficacy judgments witnessed in the quantitative domain was mainly due to the inability of a more general verbal efficacy factor to delineate satisfactorily variations in Spanish Academic Self-Efficacy.

Although clearly defined in itself, the Spanish Academic Self-Efficacy factor did not share a meaningful proportion of its variance with other verbal self-efficacy factors. In other words, students’ judgments of their academic capability in Spanish differ greatly from their efficacy perceptions in other English-based subjects such as English and U.S. history. It shows that there are more subject-specific components to each first-order self-efficacy factor than can be explained by higher order factors. Therefore, estimating subject-specific academic self-efficacy (e.g., Chemistry Academic Self-Efficacy) from self-efficacy scores in other subjects (e.g., Algebra Academic Self-Efficacy) or from more general verbal or quantitative academic self-efficacy may not be justified.

The hierarchical structure of academic self-efficacy judgments is in support of the view recently expressed by social cognitive researchers (e.g., Pajares, 1996a; Zimmerman, 1996). The specificity with which academic self-efficacy is estimated is allowed to vary depending on the researchers’ predictive and explanatory goals. Efficacy judgments can be assessed at the levels of a specific task (e.g., arithmetic-progression problems), a course (e.g., algebra), or a more general domain (e.g., mathematics). However, the specificity of self-efficacy measures should agree with the specificity of performance that the researchers wish to predict. The fact that a higher order verbal efficacy factor underlies more domain-specific factors such as English, Spanish, and U.S. History Academic Self-Efficacy does not imply that Verbal Academic Self-Efficacy should be able to predict students’ performance in English the same way English Academic Self-Efficacy does. Nor do the results reported in this article imply that task- or subject-specific efficacy perceptions can be safely inferred from related higher order self-efficacy scores. Rather, the results simply provide an empirical justification for efficacy researchers to develop and use academic self-efficacy measures at various levels of specificity that correspond to the performance of interest.

Interestingly, the second-order Verbal and Quantitative Academic Self-Efficacy factors were significantly and positively correlated. This finding stands in direct contrast to the reports that verbal and math academic self-concepts were almost always uncorrelated (Byrne & Shavelson, 1986; Marsh, 1990; Marsh, Byrne, & Shavelson, 1988; Marsh & Shavelson, 1985; Vispoel, 1995). The most potent source of efficacy information is personal mastery experiences (Ban-
dura, 1977, 1986; Schunk, 1989). Successful students in the verbal area are often successful students in the math area. It is not surprising, therefore, to find students’ efficacy perceptions in those two areas to be highly correlated. Judgments of self-efficacy are also heavily influenced by mastery criteria (i.e., being able to succeed) compared with academic self-concept assessment that tends to stress relativistic judgments of ability (i.e., being better than others; Zimmerman, 1995). The results of the present investigation contribute to the understanding of the different operations involved in academic self-concept and self-efficacy formation.

It should be noted that the generalizability of the results reported in this article inevitably depends on the selectivity of the problems used for inquiring students’ academic self-efficacy judgments. Had a different set of problems been used, different results might have emerged. Moreover, some academic tasks that students perform in each school subject are very difficult to represent as a single particularized problem. Future research on the generality of academic self-efficacy beliefs can use more divergent types of problems and tasks to probe the external validity of the present results.

Perceptions of Task Similarity and the Generality of Self-Efficacy

The present study investigated the effects of perceived similarity on the generality of academic self-efficacy judgments by using isomorphic algebra (i.e., arithmetic progression) and physics problems (i.e., constant acceleration). Previous research showed that students could not easily discern the structural parallelism in the required solution methods between the two classes of problems (Bassok & Holyoak, 1989). Students who participated in the present investigation demonstrated comparable strengths of confidence for a given pair of problems when they perceived those problems as similar. As students’ perceptions of similarity between the problems increased, the generality of their academic self-efficacy judgments between them also increased.

Although the relationships uncovered in this investigation between indicators of perceived similarity and the generality of efficacy perceptions were in the predicted direction, they were not as strong as expected in magnitude. The arithmetic progression and constant acceleration problem pairs used in the current study did not differ much along a dimension of similarity and such restrictions of range might have contributed to attenuating the observed relationship. Had more diverse problem sets been included, a stronger relationship might have been obtained.

Of course, results from the present study cannot confirm but can only suggest one of the potential mechanisms underlying the academic self-efficacy generalization. Only when experimental manipulations of similarity perception toward various academic tasks brought about the increased generalization of self-efficacy among them, can we say that we uncovered one of the major variables responsible for efficacy transfer. Also, there was much variance left to be explained at both problem and student levels after the effects of perceived similarity and class type had been controlled for. Self-efficacy is a context-dependent construct that is reciprocally related to human experience (Bandura, 1993; Zimmerman, 1995, 1996). Future research should explore the effects of other task-related as well as personal variables on the generality of academic self-efficacy beliefs.

References


Appendix A

Sample Problems for Measuring Academic Self-Efficacy

English

1. Select the most appropriate completion.

Because our supply of fossil fuel has been sadly _____, we must find _____ sources of energy.

A. stored . . . hoarded  D. increased . . . available
B. compensated . . . significant  E. depleted . . . alternate
C. exhausted . . . inefficient

2. Select the pair that best expresses a relationship similar to that in the original pair.

REBUTTAL: DISPROVE::

A. narration: summarize  D. contention: clarify
B. qualification: limit  E. annotation: define
C. refutation: conclude

3. Select the part that must be changed to make the sentence correct.

Voters justified their apathy by saying that they had no viable

A  B  C

choice because the candidates were indifferent from one an-
other. No error

D  E  F

Spanish

1. Select the most appropriate completion.

¡Qué música tan estupenda! Esta orquesta _____ muy bien.

A. juega  B. sabe  C. maneja  D. toca

2. Select the most appropriate completion.

Si _____ en el Brasil, hablaríamos portugués y no español.
A. vivamos  B. vivimos  C. vivíamos  D. viviéramos

3. Select the most appropriate completion.

Me gustó tanto la novela de Isabel Allende que _____ voy a recomendarme a mis amigos.
A. le  B. lo  C. se la  D. me la

U.S. History

1. Which of the following wars of the United States would fit the description of a war neither lost nor won?

I. The War of 1812
II. The Mexican War
III. The Spanish-American War
IV. The Second World War

A. I only  D. II and IV only
B. II only  E. III and IV only
C. I and III only

2. In the 1780's and 1790's, the United States sought the right to navigate Mississippi River in order to
A. establish a western naval force against Spain
B. satisfy New England economic interest
C. take full advantage of the Louisiana Purchase
D. provide a cheaper trade route for the farm goods of the Ohio valley
E. open lands west of the Mississippi to cotton production

3. Before 1820 most manufacturing in the United States took place in houses and small shops where
A. coal was the chief form of fuel
B. women and children often participated in the production process
C. artisans, journeymen, and apprentices worked in assembly lines to facilitate quick production of goods
D. artisans produced goods mainly for foreign export
E. artisans used tools driven by water power

Algebra

1. If □ is defined by the equation \( x \Box y = x + xy + y \) for all numbers \( x \) and \( y \), what is the value of \( z \) if \( 8 \Box z = 3 \)?

A. -5  B. \( \frac{5}{9} \)  C. \( \frac{3}{8} \)  D. \( \frac{5}{9} \)  E. 5

2. \( 2a - \frac{b}{2}\sqrt{2} = \frac{c}{25}\sqrt{2} \). In which of the following are \( a, b, \) and \( c \) arranged in descending order of value?

A. \( a, b, c \)  B. \( b, c, a \)  C. \( a, c, b \)  D. \( c, b, a \)  E. \( c, a, b \)

3. If \( \frac{x^2 + 4x + 6}{x^2 + 3x + 7} = 1 \), then \( x = \)

A. 0  B. -1  C. 1  D. \( \frac{7}{6} \)  E. 6

Geometry

1. In the figure above, radius \( OA = 6.5 \) and chord \( AC = 5 \). what is the area of \( \Delta ABC \)?

A. 28  B. 30  C. 32.5  D. 33

2. In parallelogram \( ABCD, y = 50 \). Which of the following is true?

A. \( x + y > 90 \)  C. \( x + y = 90 \)
B. \( x + y < 90 \)  D. insufficient information

3. In the figure above, \( BA = 2BC, EA = 2DE, \) and \( BE = 14 \). What is the value of \( DC \)?

A. 18  B. 20  C. 21  D. 22  E. 24
Chemistry
1. Which of the following has electrons in orbitals?
   A. Fe\(^{2+}\)  B. Cl  C. K\(^+\)  D. Cs  E. Au

2. Which of the following dissolves readily in the water to give a strongly acid solution?
   A. HCl  B. NH\(_3\)  C. N\(_2\)  D. H\(_2\)  E. CO\(_2\)

3. Which of the following combinations of particles represents an ion of net charge \(-1\) and of mass number 80?
   A. 44 neutrons, 35 protons, 36 electrons
   B. 44 neutrons, 36 protons, 35 electrons
   C. 44 neutrons, 36 protons, 36 electrons
   D. 45 neutrons, 35 protons, 35 electrons
   E. 45 neutrons, 35 protons, 36 electrons

Appendix B
Sample Algebra and Physics Problem Pairs Constructed From Bassok and Holyoak (1989)\(^B\)

Pair 1
A. What is the acceleration (= increase in speed each second) of a motorcycle if it's speed increased uniformly from 22 meters per second (22 m/s) at the beginning of the first second, to 46 m/s at the end of the 15th second?
B. During a long-term observation for his science project, Garret found that the diameter of his oak tree increases the same amount each month. If the diameter was 15 cm at the beginning of the first month, and 30 cm at the end of the 7th month, by how much does the diameter of his oak tree increase each month?

Pair 2
A. A sky-diver wearing a parachute jumped from a helicopter and falls 5.4 meters during the first second of his descent, and during each subsequent second he falls 11.8 meters farther than he fell during the preceding second. If it took him 8 seconds to reach the ground, how high above the ground was the helicopter hovering?
B. Margo Smith has a job that pays $3,000 for the first two years, with a raise of $200 at the end of every two years thereafter. What was her total income after 20 years?