AUTHOR	Bennett, Randy Elliot; Sebrechts, Marc M.
TITLE	The Accuracy of Automatic Qualitative Analyses of Constructed-Response Solutions to Algebra Word Problems. GRE Board Professional Report No. 91-03P.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-94-04
PUB DATE	Mar 94
NOTE	111p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	Algebra; Automation; Classification; College Entrance Examinations; *College Students; Computer Assisted Testing; *Ccnstructed Response; Educational Diagnosis; *Expert Systems; Higher Education; *Qualitative Research; Scoring; Test Construction; *Word Problems (Mathematics)
IDENTIFIERS	*Accuracy; *GIDE Computer Program; Graduate Record Examinations

ABSTRACT

This study evaluated expert system diagnoses of examinees' solutions to complex constructed-response algebra word problems. Problems were presented to three samples (30 college students each), each of which had taken the Graduate Record Examinations General Test. One sample took the problems in paper-and-pencil form and the other two on computer. Responses were then diagnostically analyzed by an expert system, GIDE, and by four Educational Testing Service mathematics test developers. Results were highly consistent across the samples. Human judges generally agreed in describing responses as right or wrong, but concurred at lower levels in categorizing the specific bugs they detected in incorrect solutions. The expert system agreed highly with the judges' right/wrong decisions, but less closely with bug categorizations that judges agreed on. Causes of machine-rater disagreement were identified, and suggested remedies were proposed. These results suggest that highly accurate diagnostic analysis through knowledge-based understanding of complex responses may be difficult to achieve at the fine-grained level used by GIDE. Increasing accuracy is discussed. Appendixes A, B, and C present probabilities and canonical solutions for each of the samples; and Appendixes D, E, and F contain Sample 2 judges' instructions, and Sample 2 and Sample 3 Bug Classification Scheme and Detailed Error Descriptions with Examples. Twenty-one tables present study data. (Contains 13 references.) (Author/SLD)

GRE®

RESEARCH

The Accuracy of Automatic Qualitative Analyses of Constructed-Response Solutions to Algebra Word Problems

Randy Elliot Bennett
and
Marc M. Sebrechts

March 1994

GRE Board Professional Report No. 91-03P
ETS Research Report 94-04

Educational Testing Service, Princeton, New Jersey

The Accuracy of Automatic Qualitative Analyses of Constructed-Response Solutions to Algebra Word Problems

Randy Elliot Bennett
and
Marc M. Sebrechts

GRE Board Report No. 91-03P

March 1994

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, N.J. 08541

****

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board Reports do not necessarily represent official Graduate Record Examinations Board position or policy.

****

The Graduate Record Examinations Board and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.

Abstract

This study evaluated expert system diagnoses of examinees' solutions to complex constructed-response algebra word problems. Problems were presented to three samples, each of which had taken the GRE General Test. One sample took the problems in paper-and-pencil form and the other two on computer. Responses were then diagnostically analyzed by an expert system, GIDE, and by four ETS mathematics test developers using a fine-grained categorization of error types. Results were highly consistent across the samples. Human judges agreed among themselves almost perfectly in describing responses as right or wrong but concurred at much lower levels (37% to 64% agreement) in categorizing the specific bugs they detected in incorrect solutions. The expert system agreed highly with the judges' right/wrong decisions (95% to 97% concurrence) and somewhat less closely (71% to 74%) with the bug categorizations that judges, themselves, agreed on. Seven principal causes of machine-rater disagreement were detected, most of which could be remedied by making adjustments to GIDE, modifying the test presentation interface to constrain the form of examinee solutions, and working with test developers to specify rules for automatically dealing with special cases. These results suggest that highly accurate diagnostic analysis through knowledge-based understanding of complex responses may be difficult to achieve at the fine-grained level used by GIDE. The accuracy of qualitative judgments might be increased by using a smaller set of more general diagnostic categories and by integrating information from other sources, including performance on diverse item types.

The Algebra Assessment System (Sebrechts, Bennett, & Katz, 1993) is a research tool for interactive performance assessment in graduate education. The system consists of a pool of 40 algebra word problems adapted from items in the GRE General Test's quantitative section, an interface that presents the problems and permits students to enter their step-by-step constructed responses, and GIDE (Sebrechts, LaClaire, Schooler, & Soloway, 1986), a knowledge-based program that diagnostically analyzes and numerically scores those responses.

GIDE, the first system component developed, has been the subject of several investigations. One study examined agreement of its partial-credit scores with those of human raters (Sebrechts, Bennett, & Rock, 1991). Correlations between GIDE and the mean scores taken across five raters for 12 problems ranged from .74 to .97, with a median of .88; the largest mean absolute discrepancy between GIDE and the raters was 1.2 points on a 16-point scale. A second study factor analyzed the responses of 249 examinees to ascertain the degree to which GIDE's scores related to the General Test's quantitative section, an established measure of mathematical reasoning skill (Bennett, Sebrechts, & Rock, 1991). Two highly correlated dimensions--GRE quantitative and constructed response--emerged. Along with the agreement analysis, these results suggest that GIDE can duplicate the judgments of content experts reasonably well in numerically scoring solutions to constructed-response algebra items and that these scores are consistent with those from a well-established quantitative ability measure. Finally, Bennett, Sebrechts, and Yamamoto (1991) attempted to detect diagnostically meaningful patterns across item responses for individual examinees. They found a small number of examinees consistently omitted important solution components. Most examinees, however, were inconsistent in the errors they made.

The current study examined the accuracy of GIDE's item-level qualitative analyses. These item-level analyses have several potential uses, each of which demands reasonable accuracy. First, and most important, these qualitative analyses are the basis for GIDE's partial-credit scores: Points are deducted depending upon the errors discovered in the solution process. Second, the qualitative analyses might be communicated to examinees to let them know how they did on a specific problem. This communication would serve to identify errors, help clarify the misconceptions or procedural gaps that produced those errors, and, as a consequence, help students avoid such mistakes on future problems. (In addition, such communication should enhance the credibility of automatically scored constructed-response tests by making clear why a particular score was awarded.) Finally, item-level analyses can be used as a building block for more general inferences about how examinees tend to perform across items (e.g., about what strategies an individual generally employs). These inferences might be helpful in lending greater meaning to test scores, as well as in suggesting how a given lower scoring examinee differs from a more proficient performer.

Method

Subjects

Three samples were used. The first was drawn from a pool of 249 subjects who had participated in a previous study of the accuracy of expert system scores on algebra word problems (Bennett, Sebrechts, & Rock, 1991).

This pool had been drawn from examinees who took a single form of the GRE General Test administered in June 1989 and lived near ETS field service offices. From the pool, a stratified random sample of 30 was drawn based on performance on a 12-item test of algebra word problems. Because the pool's algebra test performance was severely skewed (mean = 100.8 and standard deviation of 18.5 on a scale ranging from 0-120), the pool was stratified into eight score levels, with the bottom level oversampled to produce a more uniform distribution. The resulting group (algebra mean = 72.7, SD = 28.2) was, thus, more likely to show the full range of errors that might be encountered in the General Test population, including those at the lower proficiency levels.

The second and third samples were drawn specifically for the current study and consisted of 30 students each. GRE Programs files were searched for individuals recently taking the General Test and living in the Washington, D.C. area. Approximately 1,000 examinees were selected, placed into four groups representing quartiles in the General Test quantitative score distribution, and contacted by mail with an offer of payment to participate. From the group responding, examinees were chosen to generate a sample in which each quartile was about equally represented, with the intent of approximating the level and range of scores found in the General Test population.

Table 1 describes the three samples. Because the sizes are small, differences among samples should be interpreted cautiously. This is especially true for the categorical variables, for which there were frequently missing data. The most relevant difference appears on GRE-q, for which samples 2 and 3 were more similar to the General Test population than was sample 1. Sample 1's mean was noticeably lower and its standard deviation higher, a function of how it was selected (i.e., by oversampling examinees who scored low on a correlated mathematics test).

Instruments

Three tests were used, one for each sample. Sample 1 took Test 1, which was developed for use in a prior study (Sebrechts, Bennett, & Rock, 1991), and consisted of 12 questions based on algebra word problems from the General Test's quantitative section. Items belonged to three groups, with the questions in a group being isomorphic; that is, having the same underlying solution structure but different "cover stories." Items were based on distance = rate x time (DRT), work, and interest prototypes. The four items in each group were placed in one of four formats (see Figure 1):

Open-ended provides only the stem; the examinee must offer a correct solution.

Goal specification lists the labels for the givens and unknowns but does not provide the actual values.

Equation setup provides a set of general equations that would solve the problem.

Faulty solution consists of a variant on the correct solution that incorporates an error, although the form of the solution is otherwise generally correct.

Table 1
Demographic Data

Background Characteristic	Study Sample 1 (n=30)	Study Sample 2 (n=30)	Study Sample 3 (n=30)	1987-88 Examinee Population (n > 185,000)
General Test Performance				
Verbal mean (SD)	459 (162)	539 (99)	509 (119)	486 (122)
Quantitative mean (SD)	464 (157)	561 (121)	531 (138)	553 (139)
Analytical mean (SD)	474 (129)	582 (124)	522 (140)	529 (128)
Percentage Female	72%	63%	40%	53%
Percentage Non-White	32%	24%	30%	14%
Percentage U.S. Citizen	70%	97%	83%	81%
Graduate Major				
Social Sciences	25%	35%	24%	18%
Humanities/Arts	13%	17%	4%	11%
Life Sciences	17%	4%	8%	18%
Education	21%	13%	24%	15%
Physical Sciences	13%	4%	4%	11%
Engineering	0%	0%	8%	12%
Business	4%	0%	4%	3%
Other	8%	26%	24%	12%

Note. Population data are from Examinee and Score Trends for the GRE General Test by D. M. Wah and D. S. Robinson (Princeton, NJ, Educational Testing Service, 1990). Percentage non-White is for U.S. citizens only. Graduate major percentages are based only on examinees with decided majors. All percentages are based on the total number of examinees providing data for a given characteristic.

Figure 1
Isomorphs in Four Item Formats

Open-Ended

How many minutes will it take to fill a 2,000-cubic-centimeter tank if water flows in at the rate of 20 cubic centimeters per minute and is pumped out at the rate of 4 cubic centimeters per minute?

AIISWER: ______________________

Goal

One of two outlets of a small business is losing $500 per month while the other is making a profit of $1,750 per month. In how many months will the net profit of the small business be $35,000?

Givens
Profit from Outlet 1 = ______________________
Profit from Outlet 2 = ______________________
Target Net Profit = ______________________

Unknown
Net Monthly Profit = ______________________
= ______________________
Months to Reach Target Net Profit = ______________________

ANSWER: ______________________

Equation Setup

A specialty chemical company has patented a chemical process that involves 2 reactions. Reaction 1 generates 24 grams of molecule B per minute and reaction 2 consumes 5 grams of molecule B per minute. If 4,560 grams of molecule B are desired as a product of this process, how many minutes must it continue?

Equations That Will Provide a Solution:

Net Amount of B Per Minute = Amount. Produced by Reaction 1 + Amount. Produced by Reaction 2
Time for Desired Amount of B = Desired Amount of B/Net Amount of B Per Minute

Your Solution:

ANSWER: ______________________

Faulty Solution

$3.50 in tolls is received each minute at an automated toll booth while the rate at a booth with an operator is $2.80 each minute. How many minutes elapse before the automated booth receives $14.00 more in tolls than does the person-operated booth?

Tolls per Minute = $3.50/min + $2.80/min
Tolls per Minute = $6.30/min
Time for $14 lead = $14/$6.30 per minute
Time for $14 lead = 2.22 minutes

Your Corrected Solution:

ANSWER: ______________________

Note. From Machine-scorable complex constructed-response quantitative items: Agreement between expert system and human raters' scores (RR-91-11) by M. M. Sebrechts, R. E. Bennett, and D. A. Rock. (Princeton, NJ: Educational Testing Service, 1991).

This test, which predated development of the Algebra Assessment System interface, was administered in paper-and-pencil form in two different item orders assigned to random halves of the sample. Appendix A gives the items and their canonical solutions.

Samples 2 and 3 each took computer-based tests. Sample 2's test (Test 2) consisted of 12 items written to be isomorphs of the questions presented to sample 1 (see Appendix B). Items were assigned to one of the four formats in turn and assembled such that (1) items of a format were kept together and (2) isomorphs were separated. Two forms of the test differing in item order were administered (one form the reverse of the other), and randomly assigned to examinees.

Sample 3 took Test 3, a 12-item instrument that shared four _distance = rate x time_ questions with Test 2 (see Appendix C). Of the remaining eight items, one group of 4 was based on a _graduated rate_ problem and the other on a more difficult _distance = rate x time_ problem, designated _DRT-2_. The representation and ordering of formats were as in Test 2. In contrast with that sample, four randomly assigned presentation orders were used that better balanced the positions in which formats appeared.

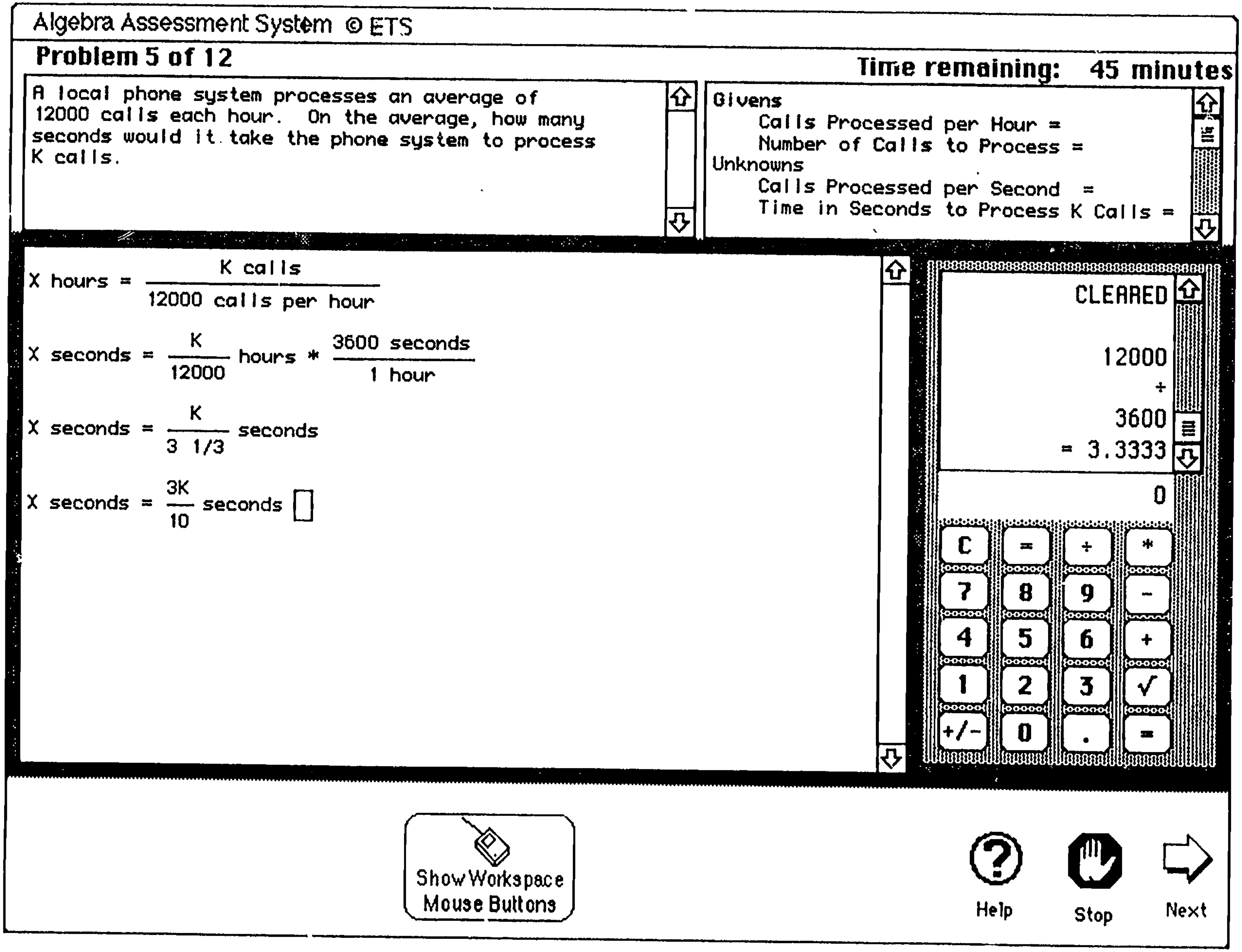
The data collection interface is depicted in Figure 2. The upper left-hand window presents a problem stem; format information is given in the window to the right. The large window below the problem stem is a workspace in which students enter equations by typing on the keyboard or by clicking on buttons using a mouse. To the right is a simple, five-function, tape calculator.

At present, the interface permits examinees to enter textual and numerical information freely. As a result, responses need to be preprocessed by a human before being passed to the expert system for analysis. Preprocessing consists primarily of removing obviously extraneous characters (e.g., semicolons in equations) and of reformulating lines so that they have only a single equal sign. (See Sebrechts, Bennett, & Katz, 1993, for a detailed description of the interface.)

Expert System

The automatic analysis mechanism was GIDE-Algebra (Sebrechts, LaClaire, Schooler, & Soloway, 1986). To analyze a response, GIDE first parses it into a standard format. It then calls upon a knowledge base. Knowledge bases are specific tc a narrow class of problems and were created in previous studies through cognitive analysis of the solutions of proficient and novice problem solvers.

For each problem, GIDE's knowledge base has a specification that identifies both the "given" information and the goals into which the problem has been decomposed, where a goal is one of several objectives to be achieved in reaching a solution (e.g., an intermediate result). To be considered correct, a solution must satisfy each goal. GIDE attempts to discover how the student solution satisfies a particular goal by testing the parsed solution



Algebra Assessment System © ETS

Problem 5 of 12

Time remaining: 45 minutes

A local phone system processes an average of 12000 calls each hour. On the average, how many seconds would it take the phone system to process K calls.

Givens
    Calls Processed per Hour =
    Number of Calls to Process =
Unknowns
    Calls Processed per Second =
    Time in Seconds to Process K Calls =

$$X \text{ hours} = \frac{K \text{ calls}}{12000 \text{ calls per hour}}$$

$$X \text{ seconds} = \frac{K}{12000} \text{ hours} * \frac{3600 \text{ seconds}}{1 \text{ hour}}$$

$$X \text{ seconds} = \frac{K}{3\ 1/3} \text{ seconds}$$

$$X \text{ seconds} = \frac{3K}{10} \text{ seconds} \ \square$$

```
CLEARED

  12000
      ÷
   3600
= 3.3333

      0
```

Figure 2. The Algebra Assessment System interface.

against a series of alternative correct plans (i.e., stereotypical procedures) drawn from its knowledge base. If no matching plan is found, GIDE attempts to discover the nature of the discrepancy by testing plans that incorporate conceptual errors commonly made in achieving that goal or that evidence more general mistakes occurring within an appropriately structured plan (e.g., in decimal placement or time conversion). When no plan, buggy or correct, can be matched, the goal is considered missing.

When the analysis of a response is completed, GIDE reports a partial-credit score (computed from the types and numbers of bugs detected) and a qualitative description of the errors found.

Procedure

GIDE's accuracy was investigated by comparing its qualitative judgments with those of content experts for each of the three samples, thus permitting replications across test delivery mode (paper-and-pencil vs. computer) and item sets. For the first sample (which took the paper-and-pencil test), two ETS mathematics test developers independently read each response, without having seen GIDE's analysis.[1] Test developers were asked to categorize examinee errors according to the classifications used by GIDE (with the addition of a category for errors not fitting GIDE's scheme). This scheme was derived from errors enumerated by test developers and from an earlier analysis of students' mistakes (Sebrechts, Bennett, & Rock, 1991). The test developers' error classifications were next compared to GIDE's analyses of the same responses to identify those instances in which the two developers were consistent in their disagreement with the expert system.

As a result of these comparisons, the knowledge bases used to analyze these items (i.e., DRT, work, interest) were revised to increase GIDE's accuracy. Ten examinees' responses from Test 2 were then used to verify these refinements and 10 from Test 3 to check the functioning of the newly built graduated rate and DRT-2 knowledge bases. These 20 subjects were eliminated from further analyses.

The responses of the remaining 30 examinees to Test 2 (i.e., sample 2) and 30 examinees to Test 3 (i.e., sample 3) were analyzed by GIDE and separately by four mathematics test developers, with each developer independently analyzing all 720 item responses (60 examinees x 12 items). As in the first analysis, the developers were asked to classify errors according to GIDE's categorization scheme, with the addition of a category for unclassifiable errors.

[1] Test developers read the original handwritten responses. GIDE analyzed transcriptions. Handwritten responses had been converted to machine-readable form as part of a prior study. The conversion was done according to rules intended to make the response amenable to machine analysis without altering its substance. Changes made to facilitate machine analysis included arranging solution elements in a linear sequence, translating each line to a syntactically correct equation (e.g., allowing only one equal sign per line), and ignoring illegible portions. After transcription, all solutions were checked for rule violations by a second transcriber.

Each of the three human scoring sessions followed the same basic fo:mat. First, the judges were given written instructions describing the task (see Appendix D for the instructions for sample 2), followed by a problem from a given class (e.g., work) and its canonical solution(s). Next, GIDE's categorization scheme for that problem class was described. This scheme took the form of a list of bugs organized into computational, specific plan, general plan, and missing goals categories, which were defined as follows:

> Computational errors involved a failure to execute a low-level operation (e.g., by inappropriately shifting a decimal, by incorrectly treating the remainder of a division as a decimal).
>
> Specific plan errors were inappropriate procedures for solving a goal linked to a particular problem class (e.g., confusing the rates for different trip segme..s in DRT problems).
>
> General plan errors suggested more universal failures to formulate procedures, with the same malformation having the potential to occur across problem contexts (e.g., dividing when multiplication is called for).
>
> Missing goals suggested the omission of a critical solution component.

The bug list was accompanied by a second list of detailed bug definitions and examples (see Appendices E and F for the lists, definitions, and examples for samples 2 and 3, respectively). Following this, the judges reviewed student solutions with bug classifications given by the session leader and then practiced applying the classification scheme to several new responses. When all judges were comfortable with the scheme, they independently analyzed all responses to that item and, in turn, to the other items from the same class. Once a problem class had been analyzed, training for the next class began and the process was repeated.

Agreement was evaluated at the solution and bug levels. At the solution level, agreement was expressed in two ways:

- The proportion agreement among experts was the number of examinee solutions for which at least a majority of raters agreed that the solution was perfect (or not), divided by the total number of examinees. Thus, a perfect solution was one in which both raters in sample 1, or three of four fo: samples 2 and 3, detected no error. A wrong solution was one in which the majority detected at least one error, regardless of whether there was consensus on the specific nature of the mistake.

- The proportion agreement between GIDE and the raters was the number of times in which GIDE and the majority of judges agreed that a response was perfect or not, divided by the number of times in which the majority of judges agreed.

At the bug level, the unit of analysis was the individual error, regardless of how many errors appeared in a single response. Agreement was expressed in five ways:

- The number of errors dete:ted by GIDE and by the raters.

- The proportion agreement among experts was calculated by dividing the number of errors for which at least a majority of raters agreed on a specific classification by the total number of errors observed by those raters. For sample 1, in which only two raters were used, the proportion agreement between raters was the number of times the two agreed divided by the total number of errors they found. This index served as a baseline against which to compare GIDE's agreement with the judges.

- The proportion agreement between GIDE and the raters was the number of times in which GIDE and the majority of judges agreed that the same error was present in a response divided by the number of times the majority of raters agreed among themselves.

- The proportion of commissions was the number of bugs the program identified that the majority of raters agreed were not present divided by the total number of bugs identified by the program.

- The proportion of omissions was the number of bugs the program failed to identify that a majority of raters agreed were present, divided by the total number of bugs agreed upon by the majority of raters. (This value is 1 - the proportion agreement between GIDE and the raters.) Omissions and commissions were not always independent in that the same disagreement might appear in both categories (e.g., if GIDE and the raters classified the same bug differently, GIDE would receive a commission error for detecting a bug not found by the raters and an omission error for not finding a bug they had identified). Thus, the proportion agreement better reflects absolute agreement, whereas the commission and omission rates give an accurate sense of the relative reasons for less-than-perfect concurrence.

Except as noted, the above analyses were conducted for each item separately and for the total set. Because each format was marked by an item from each problem class and each problem class was indicated by isomorphs in each format, the effects of format and content on accuracy were also examined.

Finally, the causes of discrepancy between GIDE and the judges in samples 2 and 3 were identified. Each discrepancy was reviewed and classified so that implications for improving GIDE's diagnostic accuracy could be drawn.

Results

Tables 2 and 3 present agreement between raters, and between GIDE and the raters, on the right/wrong status of each examinee's response in sample 1. As can be seen, the raters agreed in 99% of cases that a response was either completely correct or had one or more bugs. Similarly, GIDE agreed with the raters almost as highly, concurring on 97% of responses when taken across all items and examinees.

Tables 4 and 5 give the numbers of specific bugs detected by GIDE and the raters in sample 1. The numbers varied widely across items (from the teens to the nineties), problem formats (from 120 or so to the 190s), and problem types (from the low 100s to the 350s). Even so, the raters and GIDE detected similar numbers of bugs in each instance. Overall, in the 360 responses (30 examinees x 12 items), GIDE identified 623 bugs, rater 1 found 600 bugs, and rater 2 located 607 bugs.

Table 2
Agreement on the Right/Wrong Status of
Responses for Sample 1 by Item (n = 30)

Item	Between Raters	Between GIDE and the Raters
1 (DRT,OE)	1.00	.97
2 (%,OE)	.93	.96
3 (W,OE)	1.00	1.00
4 (%,GS)	1.00	1.00
5 (W,GS)	1.00	.93
6 (DRT,GS)	1.00	.97
7 (W.ES)	1.00	1.00
8 (DRT,ES)	1.00	.97
9 (%,ES)	1.00	.93
10 (DRT,FS)	1.00	1.00
11 (W,FS)	1.00	.93
12 (%,FS)	1.00	1.00
Total	.99	.97

Note. The _proportion agreement among experts_ was calculated by dividing the number of examinee solutions for which both raters agreed that the solution was perfect or not by the total number of examinees. The _proportion agreement between GIDE and the raters_ was the number of times in which GIDE and both judges agreed that a response was perfect or not, divided by the number of solutions on which both judges agreed.

Table 3
Agreement on the Right/Wrong Status of Responses
for Sample 1 Broken Down by Item Format
and Problem Class (n = 30)

Item	Between Raters	Between GIDE and the Raters
Format		
Open-ended	.98	.98
Goal spec.	1.00	.97
Equation setup	1.00	.97
Faulty solution	1.00	.98
Class		
Work	1.00	.97
Percent	.98	.97
DRT	1.00	.98
Total	.99	.97

Table 4
Number of Bugs Detected
for Sample 1 by Item (n = 30)

Item	GIDE	Rater 1	Rater 2
1 (DRT,OE)	83	81	82
2 (%,OE)	38	31	36
3 (W,OE)	32	26	29
4 (%,GS)	15	19	19
5 (W,GS)	15	17	15
6 (DRT,GS)	84	84	86
7 (W,ES)	28	28	29
8 (DRT,ES)	92	93	90
9 (%,ES)	38	29	29
10 (DRT,FS)	99	94	97
11 (W,FS)	37	33	32
12 (%,FS)	62	65	63
Total	623	600	607

Note. DRT = _distance = rate x time_ problem class, % = percent problem class, W = work problem class. OE = open ended, GS = goal specification, ES = equation setup, FS = faulty solution.

Table 5
Number of Bugs Detected for Sample 1 Broken Down by Item
Format and Problem Class (n = 30)

Item	GIDE	Rater 1	Rater 2
Format			
Open-ended	153	138	147
Goal spec.	114	120	120
Equation setup	158	150	148
Faulty solution	198	192	192
Class			
Work	112	104	105
Percent	153	144	147
DRT	358	352	355
Total	623	600	607

Tables 6 and 7 give the proportions of agreement and the errors of omission and commission. Overall, the raters agreed on 64% of the bugs they identified. Their agreement ranged widely across items, however, from 28% to 84%. Agreement was very similar across formats but somewhat higher for DRT (69%) than for the percent problems (57%).

GIDE agreed with 74% of the bugs the two raters agreed on. Again, agreement ranged widely across problems (from 14% to 96%). The values were somewhat higher for the faulty-solution format (80%) than for goal specification (66%) and somewhat better for the DRT (78%) than for the other problems (69%).

Over the whole item set, 31% of the bugs GIDE detected were not detected by the raters (errors of commission). This ranged from 14% (item 10) to 73% (item 5). Eight of the 12 items had commission rates of 35% or less. Commission rates were lowest for faulty solution (24%) and highest for open ended (37%). Commission rates were lower for DRT (26%) than for the other problems (38%).

GIDE failed to detect 26% of the bugs that the raters found overall (errors of omission). This ranged from 4% (item 9) to 86% (item 5). Nine of the 12 items had omission rates of 30% or less. The items with the highest errors of omission were 2 (37%), 4 (59%), and 5 (86%). Omission rates were lower for faulty solution (20%) than for goal specification (33%) and lower for DRT (22%) than for the other problems (31%).

Tables 8 through 13 give the comparable statistics for sample 2, which took a computer-based test composed of items isomorphic to those given on paper to sample 1. The results for right/wrong agreement were almost identical to those for that sample: all but perfect agreement among raters (99% of responses agreed on overall) and agreement between GIDE and the raters that was almost as high (95%). One item (#6) showed conspicuously lower machine-rater agreement than the rest but this was not replicated with its isomorph in sample 1.

About half as many bugs were found in this sample than in the prior one (which was of lower ability and lacked calculator access). As a result, the agreement estimates for individual items are somewhat unstable, making inter-item comparisons less certain. As in sample 1, GIDE and the raters generally detected similar numbers of bugs (item #6 was a notable exception), even though there was wide variation in the numbers found across items, formats, and content classes.

Over all items, the agreement among raters was considerably lower (37% with a range of 27% to 67%) than in sample 1 (64% with a range of 28% to 84%). This disparity remained when accounting for differences in the number of raters: The median agreement among all possible rater pairs was 44% for sample 2 versus 64% for the two raters in sample 1.

Even though agreement among raters was lower, GIDE's concurrence with the judges was strikingly similar: In sample 1, GIDE agreed with 74% of the diagnoses that the test developers themselves agreed on (range = 14% to 96%) versus 72% for sample 2 (range = 42% to 100%). Commission and omission rates were also comparable. Overall, the raters agreed that 31% (range = 14% to

Table 6
Agreement on Qualitative Analyses
by Item for Sample 1 (n = 30)

		Proportion Agreement		Proportion of Erroneously Detected Bugs	
Item	Mean No. of Bugs Found	Among Raters	Between GIDE and Raters	Commissions	Omissions
1 (DRT,OE)	82	.65	.73	.33	.27
2 (%,OE)	34	.68	.63	.50	.37
3 (W,OE)	28	.57	.80	.31	.20
4 (%,GS)	19	.81	.41	.53	.59
5 (W,GS)	16	.28	.14	.73	.86
6 (DRT,GS)	85	.67	.78	.26	.22
7 (W,ES)	29	.68	.74	.29	.26
8 (DRT,ES)	92	.62	.67	.34	.33
9 (%,ES)	29	.71	.96	.32	.04
10 (DRT,FS)	96	.84	.89	.14	.11
11 (W,FS)	33	.76	.71	.38	.29
12 (%,FS)	64	.42	.68	.31	.32
Total	604	.64	.74	.31	.26

Note. The mean number of bugs found is the mean of the errors detected by the two raters. The proportion agreement between raters was calculated by dividing the number of errors on which at least a majority of raters agreed by the total number of errors observed by those raters. The proportion agreement between GIDE and the raters was the number of times GIDE agreed with the raters divided by the number of times the raters agreed between themselves. The proportion of commissions was the number of bugs the program identified that both raters agreed was not present divided by the total number of bugs identified by the program. The proportion of omissions was the number of bugs the program failed to identify that both raters agreed was present divided by the total number of bugs agreed upon by both raters.

Table 7
Agreement on Qualitative Analyses for Sample 1
Broken Down by Item Format and Problem Class (n = 30)

		Proportion Agreement		Proportion of Erroneously Detected Bugs	
Item	Mean No. of Bugs Found	Between Raters	Between Raters and GIDE	Commissions	Omissions
Format					
Open-ended	143	.64	.72	.37	.28
Goal spec.	120	.62	.66	.36	.33
Equation setup	149	.65	.74	.32	.26
Faulty solution	192	.66	.80	.24	.20
Class					
Work	105	.60	.69	.38	.31
Percent	146	.57	.69	.38	.31
DRT	354	.69	.78	.26	.22
Total	604	.64	.74	.31	.26

Table 8
Agreement on the Right/Wrong Status of
Responses for Sample 2 by Item (n = 30)

Item	Among Raters	Between GIDE and the Raters
1 (DRT,OE)	1.00	.90
2 (%,OE)	1.00	.97
3 (W,OE)	1.00	.93
4 (%,GS)	1.00	1.00
5 (W,GS)	1.00	1.00
6 (DRT,GS)	.93	.79
7 (W,ES)	1.00	.93
8 (DRT,ES)	.97	.93
9 (%,ES)	1.00	1.00
10 (DRT,FS)	.97	.97
11 (W,FS)	1.00	.97
12 (%,FS)	1.00	1.00
Total	.99	.95

Note. The proportion agreement among experts was calculated by dividing the number of examinee solutions for which the majority of raters agreed that the solution was perfect or not by the total number of examinees. The proportion agreement between GIDE and the raters was the number of times in which it and the majority of judges agreed that a response was perfect or not, divided by the number of solutions on which the majority of judges agreed.

Table 9
Agreement on the Right/Wrong Status of Responses for
Sample 2 Broken Down by Item Format
and Problem Class (n = 30)

Item	Among Raters	Between GIDE and the Raters
Format		
Open-ended	1.00	.93
Goal spec.	.98	.93
Equation setup	.99	.96
Faulty solution	.99	.98
Class		
Work	1.00	.96
Percent	1.00	.99
DRT	.97	.90
Total	.99	.95

Table 10
Number of Bugs Detected
by Item for Sample 2 (n = 30)

		Raters				
Item	GIDE	Median	#1	#2	#3	#4
1 (DRT,OE)	38	39	40	36	39	39
2 (%,OE)	29	27	26	28	27	26
3 (W,OE)	14	16	16	14	19	15
4 (%,GS)	10	9	8	11	9	9
5 (W,GS)	8	9	7	8	9	9
6 (DRT,GS)	49	30	30	30	41	27
7 (W.ES)	5	2	2	2	2	2
8 (DRT,ES)	44	38	35	39	37	38
9 (%,ES)	31	30	33	30	27	30
10 (DRT,FS)	59	50	43	49	54	51
11 (W,FS)	19	19	18	12	20	19
12 (%,FS)	45	48	42	47	50	48
Total	351	310	300	306	334	313

Table 11
Number of Bugs Detected for Sample 2 Broken Down by
Item Format and Problem Class (n = 30)

		Raters				
Item	GIDE	Median	#1	#2	#3	#4
Format						
Open-ended	81	81	82	78	85	80
Goal spec.	67	47	45	49	59	45
Equation setup	80	70	70	71	66	70
Faulty solution	123	113	103	108	124	118
Class						
Work	46	44	43	36	50	45
Percent	115	113	109	116	113	113
DRT	190	155	148	154	171	155
Total	351	310	300	306	334	313

Table 12
Agreement on Qualitative Analyses
by Item for Sample 2 (n = 30)

		Proportion Agreement		Proportion of Erroneously Detected Bugs	
Item	Median No. of Bugs Found	Among Raters	Between GIDE and Raters	Commissions	Omissions
1 (DRT,OE)	39	.34	.76	.34	.24
2 (%,OE)	27	.64	.87	.21	.13
3 (W,OE)	16	.44	.42	.43	.58
4 (%,GS)	9	.30	.67	.50	.33
5 (W,GS)	9	.46	.67	.13	.33
6 (DRT,GS)	30	.35	.73	.63	.27
7 (W,ES)	2	.67	1.00	.60	.00
8 (DRT,ES)	38	.27	.65	.50	.35
9 (%,ES)	30	.54	.73	.35	.27
10 (DRT,FS)	50	.34	.60	.61	.40
11 (W,FS)	19	.40	.83	.42	.17
12 (%,FS)	48	.30	.79	.31	.21
Total	310	.37	.72	.44	.28

Note. The median number of bugs found is the median of the errors detected by the four raters. The proportion agreement among raters was calculated by dividing the number of errors on which at least a majority of raters agreed by the total number of errors observed by those raters. The proportion agreement between GIDE and the raters was the number of times GIDE agreed with the raters divided by the number of times the raters agreed between themselves. The proportion of commissions was the number of bugs the program identified that the majority of raters agreed was not present divided by the total number of bugs identified by the program. The proportion of omissions was the number of bugs the program failed to identify that the majority of raters agreed was present divided by the total number of bugs agreed upon by the majority of raters.

Table 13
Agreement on Qualitative Analyses for Sample 2
Broken Down by Item Format and Problem Class (n = 30)

		Proportion Agreement		Proportion of Erroneously Detected Bugs	
Item	Median No. of Bugs Found	Among Raters	Between Raters and GIDE	Commissions	Omissions
Format					
Open-ended	81	.43	.73	.31	.27
Goal spec.	47	.36	.71	.55	.29
Equation setup	70	.38	.71	.45	.29
Faulty solution	113	.33	.71	.47	.29
Class					
Work	44	.44	.66	.39	.34
Percent	113	.42	.78	.31	.22
DRT	155	.32	.68	.54	.32
Total	310	.37	.72	.44	.28

73%) and 44% (range = 13% to 63%) of the bugs GIDE identified were not present for samples 1 and 2, respectively; 26% (range = 4% to 86%) and 28% (range = 0% to 58%) of the bugs agreed on by the raters were not detected by GIDE.

The levels of agreement for format and content class were generally similar across the two samples. When differences among formats or content classes appeared in sample 2, none were replications of differences found in sample 1. For example, GIDE made more commission errors for goal specification (55%) than for open-ended items (31%), and for DRT (54%) than for percent (31%) problems. In sample 1, the values for these item groups were either closely similar to one another or the relationships were reversed.

Tables 14 through 19 give the results for sample 3. The findings for right/wrong agreement replicated findings from the other two samples. Interrater agreement again reached 99% and machine-rater agreement, which was previously 97% and 95%, reached 91%. In sample 3, there was somewhat more variation for machine-rater agreement across items, largely owed to the DRT class, which had slightly lower agreement than the others (85% vs. 96% for DRT-2). Relatively lower machine-rater agreement was also found for this class in sample 2, possibly because these items have more goals and, thus, longer solutions with more chances for competing interpretation.

In this sample, the number of bugs detected was similar to that observed in sample 2; in both cases GIDE generally found numbers that were comparable to those of the judges, but somewhat higher overall (409 vs. 347 for sample 3 and 351 vs. 310 for sample 2). GIDE and the raters were most alike in the number of bugs detected for the open-ended items. This result occurred in sample 2 also. Differences across samples in content class were less comparable because two of the three classes (graduated rate and DRT-2) appeared for the first time in this item set.

The agreement values among raters and between GIDE and the raters were replicated. Total agreement among raters was 46% (39% to 71%) in the current sample and 37% (27% to 67%) in sample 2. (Median agreement among all possible rater pairs was 55% in sample 3 and 44% for sample 2.) Agreement between GIDE and the raters for sample 3 was 71% (44% to 91%) and 72% for sample 2 (42% to 100%), values consistent with those from sample 1. Total commissions and omissions were also closely similar across samples: 48% and 29%, respectively, for sample 3, 44% and 28% for sample 2, and 31% and 26% for sample 1. No consistent differences in problem format or content class emerged.

To determine if agreement would be noticeably improved by broadening the error categories, we collapsed all diagnoses according to the four major groups used to organize GIDE's specific errors (i.e., computational, specific plan, general plan, missing goals). Total agreement among the raters increased considerably, from 37% (27% to 67%) to 65% (17% to 87%) in sample 2 and from 46% (39% to 71%) to 77% (19% to 100%) in sample 3. As a percentage of the bugs agreed on by the raters, machine-rater concurrence was similar to that found for the fine-grained diagnoses: 75% (50% to 100%) and 74% (57% to 91%) for the collapsed sample 2 and 3 categories, respectively, compared with 72% (42% to 100%) and 71% (44% to 91%) for the finer grained classification.

Table 14
Agreement on the Right/Wrong Status of Responses for Sample 3 by Item (n = 30)

Item	Among Raters	Between GIDE and the Raters
1 (DRT,FS)	.97	.79
2 (DRT,EQ)	1.00	.83
3 (DRT,GS)	.97	.86
4 (DRT,OE)	1.00	.90
5 (GR,OE)	1.00	.93
6 (DRT-2,OE)	.97	.90
7 (GR,GS)	.97	.93
8 (DRT-2,GS)	1.00	.97
9 (DRT-2,EQ)	.97	.97
10 (GR,EQ)	1.00	.97
11 (DRT-?,FS)	1.00	1.00
12 (GR,FS)	1.00	.90
Total	.99	.91

Table 15
Agreement on the Right/Wrong Status of Responses for Sample 3 Broken Down by Item Format and Problem Class (n = 30)

Item	Among Raters	Between GIDE and the Raters
Format		
Open-ended	.99	.91
Goal spec.	.98	.92
Equation setup	.99	.92
Faulty solution	.99	.90
Class		
DRT-2	.98	.96
Grad Rate	.99	.93
DRT	.98	.85
Total	.99	.91

Table 16
Number of Bugs Detected
for Sample 3 by Item (n = 30)

Item	GIDE	Raters Median	#1	#2	#3	#4
1 (DRT,FS)	58	51	39	50	51	54
2 (DRT,EQ)	60	52	47	51	53	53
3 (DRT,GS)	62	48	48	49	46	48
4 (DRT,OE)	45	49	47	50	52	48
5 (GR,OE)	12	9	10	7	10	7
6 (DRT-2,OE)	40	38	49	38	38	36
7 (GR,GS)	17	10	10	8	10	11
8 (DRT-2,GS)	36	32	30	30	34	33
9 (DRT-2,EQ)	16	15	15	13	14	16
10 (GR,EQ)	10	6	7	6	6	6
11 (DRT-2,FS)	36	30	26	30	31	29
12 (GR,FS)	17	11	11	11	10	10
Total	409	347	339	343	355	351

Table 17
Number of Bugs Detected for Sample 3 Broken Down by
Item Format and Problem Class (n = 30)

Item	GIDE	Raters Median	#1	#2	#3	#4
Format						
Open-ended	97	98	106	95	100	91
Goal spec.	115	89	88	87	90	92
Equation setup	86	72	69	70	73	75
Faulty solution	111	92	76	91	92	93
Class						
DRT-2	128	116	120	111	117	114
Grad Rate	56	35	38	32	36	34
DRT	225	201	181	200	202	203
Total	409	347	339	343	355	351

Table 18
Agreement on Qualitative Analyses
for Sample 3 by Item (n = 30)

Item	Median No. of Bugs Found	Proportion Agreement		Proportion of Erroneously Detected Bugs	
		Among Raters	Between GIDE and Raters	Commissions	Omissions
1 (DRT,FS)	51	.40	.76	.48	.24
2 (DRT,EQ)	52	.39	.44	.65	.56
3 (DRT,GS)	48	.48	.78	.47	.22
4 (DRT,OE)	49	.43	.66	.42	.34
5 (GR,OE)	9	.40	.67	.58	.33
6 (DRT-2,OE)	38	.41	.86	.30	.14
7 (GR,GS)	10	.44	.71	.59	.29
8 (DRT-2,GS)	32	.61	.71	.42	.29
9 (DRT-2,EQ)	15	.48	.91	.38	.09
10 (GR,EQ)	6	.67	.83	.50	.17
11 (DRT-2,FS)	30	.65	.62	.50	.38
12 (GR,FS)	11	.71	.80	.53	.20
Total	347	.46	.71	.48	.29

Note. The median number of bugs found is the median of the errors detected by the four raters. The proportion agreement among raters was calculated by dividing the number of errors on which at least a majority of raters agreed by the total number of errors observed by those raters. The proportion agreement between GIDE and the raters was the number of times GIDE agreed with the raters divided by the number of times the raters agreed between themselves. The proportion of commissions was the number of bugs the program identified that the majority of raters agreed was not present divided by the total number of bugs identified by the program. The proportion of omissions was the number of bugs the program failed to identify that the majority of raters agreed was present divided by the total number of bugs agreed upon by the majority of raters.

Table 19
Agreement on Qualitative Analyses for Sample 3
Broken Down by Item Format and Problem Class (n = 30)

Item	Median No. of Bugs Found	Proportion Agreement		Proportion of Erroneously Detected Bugs	
		Among Raters	Between Raters and GIDE	Commissions	Omissions
Format					
Open-ended	98	.42	.74	.39	.26
Goal spec.	89	.52	.75	.47	.25
Equation setup	72	.42	.58	.58	.42
Faulty solution	92	.51	.71	.50	.29
Class					
DRT-2	201	.52	.76	.40	.24
Grad Rate	35	.54	.76	.55	.24
DRT	116	.42	.66	.51	.34
Total	347	.46	.71	.48	.29

(Agreement increased, however, in terms of the total number of bugs on which GIDE agreed with the raters and as a percentage of the total number of bugs the raters detected.) Total commissions and omissions also did not change substantially in proportional terms: Commissions and omissions were 44% and 28% in sample 2 before the collapse and 38% and 25% after. The comparable figures for sample 3 were 48% and 29% before, and 42% and 26% after.

What caused the disagreements between GIDE and the judges on diagnosis using the fine-grained classification? Review of the individual discrepancies for samples 2 and 3 suggested seven specific causes and an eighth catchall category, listed in Table 20, with frequency of occurrence given in Table 21. The latter table combines sources across commission and omission errors after removing overlap between the two (both error types result when GIDE and the judges classify the same error into different bug categories). In tabulating frequency, individual discrepancies could often be attributed to more than one source; in those cases, only the most salient source for that instance was included. In addition, the values do not always represent independent sources because some sources generate multiple processing errors (e.g., a faulty parse can cause GIDE to detect several "errors" in the same correct solution).

The most frequent source (42% of instances in sample 2 and 17% in sample 3) involved discrepancies for which GIDE offered a plausible, albeit unconventional, analysis. Interestingly, the majority of discrepancies in this category occurred when the judges disagreed among themselves, with each competing analysis having some merit and GIDE providing a diagnosis comparable to at least one judge's conclusion (category 1A). Also noteworthy is that discrepancies in category 1 occurred more for DRT problems than for any other content class. This may be because DRT problems have more goals, which generate longer solutions with more chances for bugs and competing interpretations.

The second largest source, accounting for 10% and 22% of discrepancies in the two samples, stemmed from the inferences GIDE made. In some cases, these inferences were too weak. For one problem, GIDE expected the equation "5.1 + 6.83 = 11 hr 56 min," but when it encountered only "11:59," it was unable to infer that this entry was an attempt to satisfy the correct plan. At other times, GIDE made overly strong inferences. This was especially true for time manipulations in the DRT problems. For example, in specifying the number of hours, one student wrote 5.1 instead of 5.167. For the test developers this was a precision error. GIDE, however, inferred that the examinee had misrepresented time as a decimal, because 5.167 hours is equivalent to 5 hrs 10 minutes and misrepresenting 10 minutes as a decimal yields 5.10. GIDE's inference here is plausible, but very unlikely in the absence of additional evidence in the student solution.

Other inference problems resulted from the grain of analysis GIDE uses. Occasionally, the program focused too narrowly on individual values in an equation, rather than on the equation's overall value. For example, in a problem involving 12,000 calls in 1 hour, GIDE looked for calls-per-second in the form 12000/3600. When an examinee entered 120/36, GIDE interpreted the numerator and denominator as having decimal shift errors, failing to note that this expression reduces to a correct solution. Thus, looking for minor errors first, as GIDE currently does, can produce faulty matches when the more global structure is not captured (category 6C). Using too coarse a grain (6D) can

Table 20

Major Sources of Discrepancy between GIDE and Raters

1. GIDE PROVIDES A REASONABLE ANSWER: An analysis of the solution indicates that GIDE's diagnosis is a plausible interpretation.
   - A. GIDE's answer matches at least one rater.
   - B. Test developers classify bug as "other" instead of using one of GIDE's categories.
   - C. Test developers disagree among themselves and use another GIDE category.
   - D. GIDE provides a reasonable answer with which there is no test developer agreement.
2. PARSING PROBLEMS: GIDE cannot adequately parse a student's solution.
3. FORM OF EXAMINEE SOLUTION: Although the solution is parsed, GIDE cannot fully interpret it.
   - A. A textual statement is used as part of or in place of an equation.
   - B. Nonstandard or changing forms not understood by GIDE (e.g. 800 for 8:00 pm).
   - C. Values are improperly transcribed between parts of solution.
4. ORDER OF ANALYSIS: GIDE's attempt to analyzing solution lines in the order presented is inadequate.
   - A. Sequence violation: GIDE expects some goals and plan lines to be satisfied in a particular order.
   - B. Double-counting: Repeated lines in workspace and calculator are used in separate matches.
   - C. Premature capture: A line is matched, although a later line would be better.
   - D. Carry-forward: Capture of wrong lines is carried forward.
5. CRITERIA FOR MATCHING
   - A. Separate bugs are not clearly distinguishable.
   - B. Test developers use a bug type that is not on GIDE's list.
   - C. GIDE identifies less exact error than do test developers.
   - D. Numerical precision requirements are too strong or too weak.
   - E. GIDE cannot handle approximation adequately.
6. INFERENCE FAILURE: GIDE makes a wrong inference.
   - A. Inference too weak: Nonexact numerical values generate limited inference.
   - B. Inference too strong: Close numerical value is matched when it should not be.
   - C. Too fine a grain of analysis: The focus is on elements rather than equation or on one line instead of several.
   - D. Too coarse a grain of analysis: The focus is on higher order equation and the more detailed decomposition is ignored.
   - E. A mismatch on a previous goal changes the current plan match.
   - F. Implicit matching is too weak, so dependent goals are not matched.
7. PLAN ERRORS: GIDE's knowledge base is incomplete or inaccurate.
   - A. Plan missing from knowledge base.
   - B. Flawed plan: An appropriate plan is in the knowledge base but needs modification.
   - C. Plan sequence: A less adequate plan is triggered before one that would produce a better match.
   - D. Plan interaction failure.
8. UN. .PLAINED DISCREPANCY: A source could not be clearly identified.

Table 21
Percentage of Discrepancieь Attributable to Different Sources in Approximate Order by Frequency

Source	Sample 2	Sample 3
GIDE Provides Reasonable Answer	42%	17%
Inference Failure	10%	22%
Form of Solution	12%	13%
Order of Analysis	8%	13%
Parsing Problem	7%	16%
Plan Errors	9%	10%
Criteria for Matching	10%	6%
Unexplained Discrepancy	2%	3%
Total Number of Discrepancies	214	241

also produce difficulties, as when examinees break out small steps that GIDE expects will be combined. This was especially the case with DRT problems involving time, in which several examinees converted hours to minutes, worked subparts of the problems, and then converted back to hours. GIDE did not anticipate this level of decomposition.

The third major source concerned the form of the solution. This source accounted for 12% and 13% of the discrepancies in the two samples. In many cases students used text, which GIDE cannot interpret, in place of equations. In other cases, examinees produced parsable expressions that were not well formed and, therefore, apt to be misunderstood. For example, one examinee wrote 1 hr = $116 - $20, probably to represent that the first hour of a service was charged at a rate of $20 subtracted from a total of $116. GIDE, however, interpreted this input as an equation, and therefore determined that, although a correct expression ($116 - $20) was present, a calculation error had been made (116 - 20 = 96, not 1). Examinees were also inconsistent in the symbolic forms they used. In problems using time, some switched between decimal (e.g., 6.30) and clock (e.g., 6:30) notations. This was sometimes an error and sometimes shorthand, a difference GIDE cannot detect (it always interprets the decimal as a portion of a unit and not as a colon substitute). Other form-related processing faults resulted from errors in transcribing values from the problem statement or the calculator to the workspace. GIDE does check for numbers that are numerically close to expected values, so that although it can properly interpret transcription mistakes that fall within certain bounds, it does not have clear rules for all cases.

The fourth source was associated with the sequence in which solution components were analyzed and accounted for 8% and 13% of discrepancies in the samples. GIDE's knowledge base orders solution steps according to the sequence in which they would normally be generated, so GIDE looks for that sequence in the examinee's solution. Unfortunately, lines entered in the Algebra Assessment System's workspace are stored together and precede all work the examinee did on the calculator, with no indication of how the calculator and workspace productions relate. Thus, the stored solution does not reflect the order generated by the student, in which calculator and workspace productions are interleaved. As a consequence, GIDE's processing sometimes fails, because goals can occur out of the expected sequence. A second mal-effect of separating calculator and workspace information is that GIDE often counts related calculator and workspace input as separate, assigning lines belonging to the same plan to separate goals.

Sequence underlies two other processing faults. In some cases, GIDE finds a line that fits its expectation, although a subsequent line--which it then ignores--would provide a better match. Second, processing errors are carried forward. GIDE routinely uses values it has associated with prior goals to identify plans that satisfy subsequent goals. Thus, if GIDE's analysis mistakenly indicates that the examinee has assigned a value of "1" to "116 - 20" (as above), it will search for lines that use "1" as an operand or result. This is a useful feature when the value was intended as a calculation result. However, when GIDE's interpretation is not what the examinee intended, processing can fail.

The fifth major source resulted from GIDE's inability to parse an examinee's solution into an interpretable form. This category accounted for only a few discrepancies in sample 2 (7%) but a more sizable portion in sample 3 (16%). Parsing errors tended to occur when an examinee used a nonstandard representation or included periods or other punctuation in an equation that were missed in preprocessing. For the DRT-2 problems, GIDE did not reliably parse all uses of the constant into a form it could recognize. Because it operates on the parsed solution, its diagnoses were consequently based on productions different from the ones the students provided.

Knowledge base problems constituted the sixth category of discrepancy (9% in sample 1 and 10% in sample 2). In some instances GIDE was missing plans that were required to interpret an examinee's solution. These instances were due to a small number of unanticipated strategies, many of which were unusual paths to solution. For example, in solving a problem involving the rate of electrical components produced per hour, an examinee converted to minutes and then made errors with the converted units. Because the problem did not require a conversion, GIDE did not have the plans needed to interpret those errors. In another case, the examinee used "D" to represent a value given in terms of "C" dollars.

The last major identifiable source resulted from differences in the matching criteria used by the judges and GIDE. This source accounted for 10% and 6% of instances. In some cases, separate bugs were not clearly distinguishable. In other instances, the judges found bugs that were not part of GIDE's knowledge base or that were more specific than the bugs identified by GIDE. There were also differences in the numerical precision used by GIDE and the test developers. In general, examinees were to compute solutions to the nearest minute for time problems and to the nearest hundredth for all other items. GIDE sometimes took this requirement too literally. For example, a problem concerning fund raising during a television break had an exact answer of 6:23.13 p.m. Although the nearest minute was technically 6:23, an answer of 6:24 is reasonable because the next minute had already started. GIDE, however, tried to match the nearest minute, so 6:24 was inappropriately recorded as an examinee error. A final difference concerned approximation; GIDE required values to be within certain well-specified numerical constraints and could not handle values that were close to--but outside of--these limits.

Discussion

This study assessed the accuracy of an expert system's diagnoses of students' constructed-response solutions to algebra word problems. Accuracy in qualitative diagnosis is important if we are to build computer-delivered tests and self-assessment tools that present and score complex constructed-response problems. For such devices, qualitative diagnoses can serve as the basic elements for computing partial-credit scores and modeling performance across items, for the audit trail needed to explain how those scores were arrived at, and for assistive feedback to the examinee.

In this study, human judges agreed among themselves about whether errors were present in a solution but to only a limited degree on the specific nature of those errors: 64% of the total bugs identified in sample 1, 37% of the total in sample 2, and 46% in sample 3. Agreement increased when the diagnostic categories were collapsed to form a more general classification scheme. These results suggest that it is difficult to reliably characterize errors in solutions to complex constructed-response algebra problems, at least at the fine-grained level used by GIDE. Reliability is undoubtedly lowered by the fact that student intentions are often unclear, forcing judges to rely more upon their own differing intuitions in determining how an error might best be described.

GIDE agreed very closely with the judges in characterizing responses as right or wrong. On classifying specific errors, GIDE concurred with the judges somewhat less frequently, duplicating what they, themselves, agreed on in roughly 7 of 10 instances overall. When disagreements did occur, they were more frequently errors of commission than omission (i.e., of GIDE detecting bugs that the judges agreed did not exist). These overall results belied considerable variation across items, some of which likely derived from unstable estimates associated with small samples. However, some differences in agreement levels were associated with real disparities among items in the character of solutions and in how those solutions were processed. In general, the above findings were consistent across three examinee samples that differed in the items encountered and the test delivery mode (paper-and-pencil vs. computer). Finally, a detailed analysis of discrepancies for the two computer-based samples indicated that a substantial portion were instances in which GIDE's descriptions were reasonable ones.

These results differ from analyses of the rater reliability of partial-credit scores. Using the same items as in sample 1, Sebrechts, Bennett, and Rock (1991) found (1) judges to agree highly among themselves and (2) GIDE to agree acceptably well with the judges. The contrasting findings are likely owed to the fact that the rules for computing partial-credit scores required that raters make only gross distinctions among errors. While GIDE based its partial-credit scores on its fine-grained diagnostic judgments, the mapping of judgments to score categories was many-to-one, such that diagnostic judgments could be confused without changing a response's partial-credit score.

The finding of greater accuracy for partial-credit scores than qualitative diagnoses was also noted in the domain of computer science. Here, Braun, Bennett, Frye, and Soloway (1990) used an expert system called MicroPROUST to analyze open-ended responses to computer programming problems, with the partial-credit scores again being generated from the diagnostic analysis. MicroPROUST agreed with a human rater on the qualitative interpretation of bugs for only 54% of the 706 errors discovered. However, its partial-credit scores correlated .86 with the scores assigned by that same rater.

Even though expert systems can score responses acceptably well using only modestly precise qualitative diagnoses, better diagnosis may be required if assistive feedback and audit trails are to function effectively, and if the precision of partial-credit scores is to be improved. In our study, diagnostic disagreements between GIDE and the judges emanated from six specific sources. To what extent are these problems correctable?

Inference failures, the most prominent source, pose difficult, but not insurmountable, problems. One such failure results from focusing narrowly on individual values rather than on an equation's overall value (category 6C in Table 20). This failure might be prevented by generalizing existing mechanisms. One such mechanism ignores minor errors if an acceptable expression is found leading to a correct value. Another mechanism checks for the specific form of the equation and a third for the value of the expression. These latter two mechanisms might be combined to detect alternative expression forms when no correct result is given. Thus, in finding 120/36 as a representation of 12,000/3,600, checking the equation form would show decimal errors but checking the expression's value would reveal a correct result.

A second inference failure occurs when particular values in a problem solution can come from multiple sources (including common errors) (category 6B). Because GIDE uses numerical values to help determine what the student is trying to do, it will sometimes make wrong inferences. This failure might be alleviated by constructing problems with clearly distinguishable values. The problems we employed were adapted from tests that did not permit calculator use, so values were generally chosen from a relatively small set of easy-to-work-with numbers. Because this numerical pool is limited, there is opportunity for confusion. As test items are written to be solved with calculators, the range of usable numbers should increase, making it easier for GIDE to differentiate among values in the responses.

Processing errors due to unusually formed but parsable solutions are also prominent and difficult to address. For example, students mix text and numerical values frequently (category 3A). Whereas simpler text strings might be interpreted by an improved parser, a more general facility would be required for understanding English expressions (e.g., a semantic pattern-matching capability like that of Kaplan, 1992). The switching of forms of representation (category 3B) also is challenging because there is no general solution for distinguishing when a value is the erroneous result of an equation or a notational convenience (as in 1 = 116 - 20). One possible strategy is for GIDE to withhold the assignment of an error until a final answer is provided. If the answer is correct, then the associated errors can be eliminated. This tack is now taken within certain goals, but extending the approach across goals would be more complicated: Because the analysis of the current goal depends on the results from prior goals, introducing a backwards dependency could create an unresolvable circular reference.

A better solution might be to design the interface so that only interpretable equations were accepted. Computational errors would be signaled to the examinee immediately. Examinees could write whatever they chose but any equations constructed would have to be well formed. Similarly, transcription error (category 3C) could be addressed by allowing examinees to transfer values directly from the problem stem, to and from the calculator, and to and from the workspace, rather than requiring that numbers be retyped.

Order-of-analysis difficulties could be handled with a few modifications to GIDE or to the interface. First, GIDE could be programmed to check first for the correct end result. Thus, if a student made an error and then later corrected it, GIDE would not wrongly trap the buggy attempt. This would reduce premature capture (category 4C) and the carrying forward of incorrect values (category 4D). Problems with expected sequencing (category 4A) and double-counting (category 4B) could be solved largely by integrating the information from the calculator and workspace into a single entry and storage sequence.

Parsing difficulties can be reduced substantially by changing the interface to constrain the response (e.g., to prohibit punctuation in expressions or more than one equal sign per line). Other difficulties, including processing the various ways in which students formulate equations with constants, would require new parsing routines.

Discrepancies due to missing plans can be reduced by adding to GIDE's knowledge base. Missing plans should decrease, reaching a lower asymptote, as the number of students used to build and refine the knowledge base grows.

Differences in the criteria employed for matching can likewise be dealt with by modifications to the knowledge bases. Nondistinguishable bugs can be merged (category 5A) and new bugs identified by judges can be added (category 5B). Numerical precision problems (category 5D) could also be addressed in this way. For example, rounding up to the next minute in time problems can be added as acceptable. In order to deal with approximation (category 5E), procedures for determining what counts as a close answer would have to be specified. Developing rules may be difficult because of the potential for confusing different approximate values in a given problem with values that stem from common errors. GIDE already has a mechanism for accepting some deviations from expected value and rejecting others. Those mechanisms could be extended if an appropriate set of approximation rules can be defined.

In sum, most discrepancy sources can be addressed, increasing agreement and reducing variation in processing accuracy. These results can be achieved by (1) adjusting GIDE's knowledge base and inference engine, (2) modifying the test presentation interface to constrain the solution form, and (3) working with test developers to specify rules for automatically processing unusual cases. Problems that will require greater effort include interpreting English language expressions and dealing with such inconsistent behavior as switching the meaning of notation. Fortunately, these situations appear rarely and can be reduced somewhat through interface modifications.

What does this study imply for the GRE program? The major implication is that consistent, highly accurate diagnostic analysis through knowledge-based understanding of complex constructed responses may be difficult to achieve at the fine-grained level used by GIDE. One strategy for obtaining greater analytic accuracy may be to use a coarser diagnostic classification. Although the general categories employed by GIDE (computation, general plan, specific plan, missing goals) seemed a reasonable possibility, using this scheme did not substantially improve GIDE's concurrence with the raters (although it did improve the raters' agreement among themselves). The problem-solving phases suggested by Mayer, Larkin, and Kadane (1984) (translation, understanding, planning, execution) might also be tried. For many purposes, such a gross categorization may be sufficient.

In addition to using a coarser classification, greater accuracy might be had by combining GIDE's analysis with other information. One example might be asking the student for clarification when inconsistent or uninterpretable behavior is encountered. Another example is performance on other tasks, so that multiple methods are used to make diagnostic judgments. In such a scheme, probabilistic models like inference networks might be employed to connect performance on the various items to a small number of qualitative proficiency descriptions (Mislevy, 1993). This idea might deemphasize bug information, which may have little direct instructional value in some circumstances (Sleeman, Kelly, Martinak, Ward, & Moore, 1989), and instead use that information behind-the-scenes to help generate descriptions about the skills examinees possess.

A realization of this idea might be a computer-based assessment system containing multiple-choice items, complex constructed-response tasks scorable by GIDE, and tasks that ask the examinee to classify (but not solve) problems on the basis of mathematical structure. Each task type might provide information mapped to one or more elements of Mayer et al.'s framework: the classification tasks to problem understanding, the complex constructed-responses to understanding, planning, and execution, and the multiple-choice tasks to translation, planning and execution. (Bejar, Embretson, & Mayer, 1987, give examples of multiple-choice items built around this model.) Such a mapping might reveal that some examinees are more adept at understanding problems than they are at planning and executing their solutions. Other examinees may be able to construct an adequate representation only after attempting to execute several inappropriate plans. Such cognitive differences have obvious instructional implications but also might afford predictive information beyond that contained in conventional quantitative tests like the GRE General Test quantitative section.

The Algebra Assessment System (Sebrechts, Bennett, & Katz, 1993), of which GIDE is a part, offers an existing structure to test this notion. Necessary extensions would involve adding the capability to present and record responses to multiple-choice items, integrating the problem-classification response type now under development (Bennett, 1992), and adding the inference network. Developing the network would require research on several topics. For one, we need a deeper understanding of the relationship among the many specific errors that GIDE detects so that these errors can be organized into more general categories that facilitate stable qualitative characterizations. (See Sebrechts, Enright, Bennett, & Martin, 1993, for an initial step in this direction.) Second, we must refine our theory of item formats' cognitive demands so we can better understand how formats might be used in describing examinee performance. Finally, we will need to test the resulting task theory and error categorization, making further refinements and empirical analyses iteratively until we arrive at a functional, cognitively driven measur ment model.

References

Bejar, I. I., Embretson, S., & Mayer, R. E. (1987). Cognitive psychology and the SAT: A review of some implications (RR-87-28). Princeton, NJ: Educational Testing Service.

Bennett, R. E. (1992). A new task type for measuring the representational component of quantitative proficiency, GRE Research Proposal.

Bennett, R. E., Sebrechts, M. M., & Rock, D. A. (1991). The convergent validity of expert system scores for complex constructed-response quantitative items (RR-91-12). Princeton, NJ: Educational Testing Service. (Also Applied Psychological Measurement, 1991, 15, 227-239)

Bennett, R. E., Sebrechts, M. M., & Yamamoto, K. (1991). Fitting new measurement models to GRE General Test constructed-response item data (RR-91-60). Princeton, NJ: Educational Testing Service.

Braun, H. I., Bennett, R. E., Frye, D, & Soloway, E. (1990). Scoring constructed responses using expert systems. Journal of Educational Measurement, 27, 93-108.

Kaplan, R. M. (1992). Using a trainable pattern-directed computer program to score natural language item responses (RR-91-31). Princeton, NJ: Educational Testing Service.

Mayer, R. E., Larkin, J. H., & Kadane, J. B. (1984). A cognitive analysis of mathematical problem-solving ability. In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (pp. 231-273). Hillsdale, NJ: Erlbaum.

Mislevy, R. J. (1993). A framework for studying differences between multiple-choice and free-response test items. In R. E. Bennett & W. C. Ward (Eds.), Construction vs. choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment (pp. 75-106). Hillsdale, NJ: Erlbaum.

Sebrechts, M. M., Bennett, R. E., & Katz, I. R. (1993). A research platform for interactive performance assessment in graduate education (RR-93-08). Princeton, NJ: Educational Testing Service.

Sebrechts, M. M., Bennett, R. E., & Rock, D. A. (1991). Machine-scorable complex constructed-response items: Agreement between expert system and human raters' scores (RR-91-11). Princeton, NJ: Educational Testing Service. (Also Journal of Applied Psychology, 1991, 76, 856-862.)

Sebrechts, M. M., Enright, M., Bennett, R. E., & Martin, K. (1993). Toward a cognitive basis for quantitative ability measures (RR-93-22). Princeton, NJ: Educational Testing Service.

Sebrechts, M. M., LaClaire, L., Schooler, L. J., & Soloway, E. (1986). Toward generalized intention-based diagnosis: GIDE. In R. C. Ryan (Ed.), Proceedings of the 7th National Educational Computing Conference (pp. 237-242). Eugene, OR: International Council on Computers in Education.

Sleeman, D., Kelly, A. E., Martinak, R., Ward, R. D., & Moore, J. L. (1989). Studies of diagnosis and remediation with high school algebra students. Cognitive Science, 13, 551-568.

Appendix A

Sample 1 Problems and Canonical Solutions

Work-1

How many minutes will it take to fill a 2,000 cubic centimeter tank if water flows in at the rate of 20 cubic centimeters per minute and is simultaneously pumped out at the rate of 4 cubic centimeters per minute?

Canonical Solution

Net Rate of Filling = 20 cc per minute - 4 cc per minute
Net Rate of Filling = 16 cc per minute
Time to Complete Filling = 2000 cc / 16 cc per minute
Time to Complete Filling = 125 minutes

Work-2

One of the two outlets of a small business is losing $500 per month while the other is making a profit of $1,750 per month. In how many months will the net profit of the small business be $35,000? (GS)

Canonical Solution

Net Monthly Profit = $1,750 per month - $500 per month
Net Monthly Profit = $1,250 per month
Months to Reach Target Net Profit = $35,000 / $1,250 per month
Months to Reach Target Net Profit = 28 months

Work-3

A specialty chemical company has patented a chemical process that involves 2 reactions. Reaction 1 generates 24 grams of molecule B per minute and reaction 2 consumes 5 grams of molecule B per minute. If 4,560 grams of molecule B are desired as a product of this process, how many minutes must it continue?

Canonical Solution

Net Amount of B per Minute = 24 grams per minute - 5 grams per minute
Net Amount of B per Minute = 19 grams per minute
Time for Desired Amount of B = 4,560 grams / 19 grams per minute
Time for Desired Amount of B = 240 minutes

Work-4

$3.50 in tolls is received each minute at an automated toll booth while the rate at a booth with an operator is $2.80 each minute. How many minutes elapse before the automated booth receives $14.00 more in tolls than does the person-operated booth?

Canonical Solution

Difference in Toll Booth Rates = $3.50 per minute - $2.80 per minute
Difference in Toll Booth Rates = $0.70 per minute
Time for Desired Lead = $14 / $0.70 per minute
Time for Desired Lead = 20 minutes

Interest-1

Money invested in a certain fund .arns the same dividend each year, which is 5 percent of the original amount invested. In h . many years will $750 invested in the fund earn total dividends equal to this amount invested?

Canonical Solutions

#1
Yearly Dividend = $750 * 5% per year
Yearly Dividend = $37.50 per year
Years Needed for Dividends to Equal Original Investment = $750 / $37.50 per year
Years Needed for Dividends to Equal Original Investment = 20 years

#2
100% Dividend = 5% Dividend per Year * X Years
X Years = 100% Dividend / 5% Dividend per Year
X Years = 20 Years

Interest-2

On every $150 load of cement it delivers to a construction site, Acme Cement Company earns a 4 percent profit. How many loads must it deliver to the site to earn $150 in profit?

Canonical Solutions

#1
Profit per Load = $150 * 4% per load
Profit per Load = $6 per load
Loads Needed for Target Profit = $150 / $6 per load
Loads Needed for Target Profit = 25 loads

#2
100% Profit = 4% Profit per Load * X Loads
X Loads = 100% Profit / 4% Profit per Load
X Loads = 25 Loads

Interest-3

A graphics designer earns 2% of a $1500 yearly bonus for each shift of overtime she works. How many shifts of overtime must she work to earn the equivalent of the entire yearly bonus?

Canonical Solutions

#1
Amount Earned for Each Overtime Shift = $1500 * 2% per shift
Amount Earned for Each Overtime Shift = $30 per shift
Number of Shifts for Yearly Bonus = $1500 / $30 per shift
Number of Shifts for Yearly Bonus = 50 shifts

#2
100% of Amount Earned = 2% of Amount Earned per Shift * X Shifts
X Shifts = 100% of Amount Earned / 2% of Amount Earned per Shift
X Shifts = 50 Shifts

Interest-4

The active ingredient is 0.25 percent of a 3-ounce dose of a certain cold remedy. What is the number of doses a patient must take before receiving 3 ounces of the active ingredient?

Canonical Solutions

#1
Ounces of Active Ingredient per Dose = 0.25 Percent per Dose * 3 ounces
Ounces of Active Ingredient per Dose = 0.0075 Ounces per Dose
Number of Doses Required = 3 Ounces / 0.0075 Ounces per Dose
Number of Doses Required = 400 Doses

#2
100% Active Ingredient = 0.25% Active Ingredient per Dose * X Doses
X Doses = 100% Active Ingredient / 0.25% Active Ingr.dient per Dose
X Doses = 400 Doses

Distance = Rate x Time-1

On a 600-hundred mile motor trip, Bill averaged 45 miles per hour for the first 285 miles and 50 miles per hour for the remainder of the trip. If he started at 7:00 a.m. and drove nonstop, at what time, to the nearest minute, did he finish the trip?

Canonical Solution

Time for First Part of Trip = 285 miles / 45 mph
Time for First Part of Trip = 6 1/3 hours
Distance of Second Part of Trip = 600 miles - 285 miles
Distance of Second Part of Trip = 315 miles
Time for Second Part of Trip = 315 miles / 50 mph
Time for Second Part of Trip = 6 3/10 hours
Time for Total Trip = 6 1/3 hours + 6 3/10 hours
Time for Total Trip = 12 19/30 hours
Time for Total Trip = 12 hours 38 minutes
Ending Time of Trip = 7:00 a.m. + 12 hours 38 minutes
Ending Time of Trip = 7:38 p.m.

Distance = Rate x Time-2

Workers started filling a 2,400 gallon tank through a single hose at 5:30 a.m. The first 800 gallons flowed into the tank at the rate of 75 gallons per hour through a twisted hose. Then, the twist was discovered and eliminated and the rest of the tank is being filled at the rate of 250 gallons per hour. At what time, to the nearest minute, will the filling of the tank be finished?

Canonical Solution

Filling Time 1 = 800 gallons / 75 gallons per hour
Filling Time 1 = 10.67 hours
Filling Amount 2 = 2,400 gallons - 800 gallons
Filling Amount 2= 1,600 gallons
Filling Time 2 = 1,600 gallons / 250 gallons per hour
Filling Time 2 = 6.4 hours
Total Filling Time = 10.67 hours + 6.4 hours
Total Filling Time = 17.07 hours
Total Filling Time = 17 hours 4 minutes
Ending Time for Filling = 5:30 a.m. + 17 hours 4 minutes
Ending Time for Filling = 10:34 p.m.

Distance = Rate x Time-3

A secretary typed the first 1,960 characters of a 6,424-character report at the rate of 105 characters per minute and is typing the rest of the report at the rate of 90 characters per minute. If the secretary began typing the report at 9:30 a.m. and types without interruption, at what time, to the nearest minute, will the secretary finish?

Canonical Solution

Typing Time for First Set = 1,960 characters / 105 characters per minute
Typing Time for First Set = 18.67 minutes
Number of Characters in Second Set = 6,424 characters - 1,960 characters
Number of Characters in Second Set = 4,464 characters
Typing Time for Second Set = 4,464 characters / 90 characters per minute
Typing Time for Second Set = 49.6 minutes
Total Typing Time = 18.67 minutes + 49.6 minutes
Total Typing Time = 68.27 minutes
Ending Time for Typing = 9:30 a.m. + 1 hour 8 minutes
Ending Time for Typing = 10:38 a.m.

Distance = Rate x Time-4

A certain snowplow cleared the 10.4 mile straight part of a 16.7 mile highway at the rate of 3.2 miles per hour and cleared the remaining winding portion of the highway at the rate of 2.7 miles per hour. If the snowplow started at 11:15 a.m. and ran continuously, at what time, to the nearest minute, did it finish clearing the highway?

Canonical Solution

Time for Plowing Straight Part = 10.4 miles / 3.2 mph
Time for Plowing Straight Part = 3.25 hours
Distance of Winding Part = 16.7 miles - 10.4 miles
Distance of Winding Part = 6.3 miles
Time for Plowing Winding Part = 6.3 miles / 2.7 mph
Time for Plowing Winding Part = 2.33 hours
Time for Total Plowing = 2.33 hours + 3.25 hours
Time for Total Plowing = 5 hours 35 minutes
Ending Time of Plowing = 11:15 a.m. + 5 hours 35 minutes
Ending Time of Plowing = 4:50 p.m.

Appendix B

Sample 2 Problems and Canonical Solutions

Problem Statements and Correct Solutions

The following pages contain 12 problems, 4 isomorphs of each of 3 problem types (WORK, PERCENT, DISTANCE=RATExTIME). Each problem stem is identified by its first letter as belonging to one of these three types. Each problem also has a unique number from 1 to 12.

The lines in each solution below are numbered in a way that corresponds to the "Canonical Solutions" used with the "Detailed Error Descriptions."

Lines in the correct solutions are grouped into goals; a blank line separates each goal. When there is an "A" and "B" version of a line, it means that either form is acceptable.

W-03. If tickets for the evening performance of an event sell at the rate of 125 per hour, while tickets for the matinee performance sell at the rate of 40 per hour, after how many hours do ticket sales for the evening performance exceed ticket sales for the matinee performance by 1,020 tickets?

Correct solution A

1. Difference in Rate of Ticket Sales = 125 tickets per hour - 40 tickets per hour
2. Difference in Rate of Ticket Sales = 85 tickets per hour

3. Time for Evening Ticket Sales to Exceed Matinee Ticket Sales by Desired Amount =
   1,020 tickets / 85 tickets per hour
4. Time for Evening Ticket Sales to Exceed Matinee Ticket Sales by Desired Amount = 12 hours

Correct solution B

1. Time for Evening Ticket Sales = 1020 tickets / 125 tickets per hour
2. Time for Evening Ticket Sales = 8.16 hours
3. Time for Matinee Ticket Sales = 1,020 tickets / 40 tickets per hour
4. Time for Matinee Ticket Sales = 25.5 hours

5. Time for Net Ticket Sales = 1 unit / (1 unit/8.16 hours - 1 unit/25.5 hours)
6. Time for Net Ticket Sales = 1 unit / (0.1225 hours - 0.0392 hours)
7. Time for Net Ticket Sales = 12 hours

W-05. If the Smith household uses 250 gallons of water per day, whereas the Russell household uses 175 gallons per day, how many days elapse before the Smiths use 3,750 more gallons of water than the Russells?

Correct solution A

1. Difference in Amount of Water Used per Day = 250 gallons per day - 175 gallons per day
2. Difference in Amount of Water Used per Day = 75 gallons per day

3. Time to Reach Target Difference in Water Use = 3,750 gallons / 75 gallons per day
4. Time to Reach Target Difference in Water Use = 50 days

Correct solution B

1. Time for Smith Water Use = 3,750 gallons / 250 gallons per day
2. Time for Smith Water Use = 15 days
3. Time for Russell Water Use = 3,750 gallons / 175 gallons per day
4. Time for Russell Water Use = 21.43 days

5. Time for Net Water Use = 1 unit / (1 unit/15 days - 1 unit/21.43 days)
6. Tim: for Net Water Use = 1 unit / (0.0667 days - 0.0467 days)
7. Time for Net Water Use = 50 days

W-07. If money is deposited into a soda machine at the rate of $12.00 per hour and is returned as change at the rate of $3.60 per hour, how many hours will it take the soda machine to accumulate $126.00?

Correct solution A

1. Amount Accumulated per Hour = $12.00 per hour - $3.60 per hour
2. Amount Accumulated per Hour = $8.40 per hour

3. Time to Accumulate Desired Amount of Money = $126.00 / $8.40 per hour
4. Time to Accumulate Desired Amount of Money = 15 hours

Correct solution B

1. Time for Deposits = $126.00 / $12.00 per hour
2. Time for Deposits = 10.5 hours
3. Time for Change Returned = $126.00 / $3.60 per hour
4. Time for Change Return = 35 hours

5. Time for Net Accumulation = 1 unit / (1 unit/10.5 hours - 1 unit/35 hours)
6. Time for Net Accumulation = 1 unit / (0.0952 hours - 0.0286 hours)
7. Time for Net Accumulation = 15 hours

W-11. A machine produces 35 bottles per minute, of which 3 are defective. How many minutes does the machine require to produce 8,000 nondefective bottles?

Correct solution A

1. Nondefective Bottles per Minute = 35 bottles per minute - 3 bottles per minute
2. Nondefective Bottles per Minute = 32 bottles per minute

3. Time for Desired Nondefective Bottles = 8,000 bottles / 32 bottles per minute
4. Time for Desired Nondefective Bottles = 250 minutes

Correct solution B

1. Time for All Bottles = 8,000 bottles / 35 bottles per minute
2. Time for All Bottles = 228.57 minutes
3. Time for Defective Bottles = 8,000 bottles / 3 bottles per minute
4. Time for Defective Bottles = 2666.67 minutes

5. Time for Nondefective Bottles = 1 unit / (1 unit/228.57 minutes - 1 unit/2666.67 minutes)
6. Time for Nondefective Bottles = 1 unit / (0.0044 minute - 0.0004 minute)
7. Time for Nondefective Bottles = 250 minutes

DRT-01. A fire is burning in a forest of several thousand acres. Firefighters are attempting to confine the fire to one region covering 705 acres, and plan to light a backfire at the moment the fire reaches the edge of that region, after having burned the entire region. The first 459 acres burned at a rate of 90 acres per hour. The wind then diminished, and the fire is now burning at the rate of 36 acres per hour. If the fire started at 10 a.m. and wind conditions do not change, at what time, to the nearest minute, should the backfire be lit?

Correct solution

1. Burning Time1 = 459 acres / 90 acres per hour
2. Burning Time1 = 5.1 hours

3. Acres Amount2 = 705 acres - 459 acres
4. Acres Amount2 = 246 acres

5. Burning Time2 = 246 acres / 36 acres per hour
6. Burning Time2 = 6.83 hours

7. Total Burning Time = 5.1 hours + 6.83 hours
8. Total Burning Time = 11 hours 56 minutes

9. Time Backfire Should Be Lit = 10 a.m. + 11 hours 56 minutes
10. Time Backfire Should Be Lit = 9:56 p.m.

DRT-06. During a break in its program at 6:15 p.m., a TV station began a fund-raising campaign that has a goal of raising $13,900 through viewer solicitation. It raised the first $7,000 at the rate of $1,200 per minute. When a special appeal began, pledges came in at the rate of $3,000 per minute. If funds continue to come in at the same rate as during the special appeal, until what time, to the nearest minute, must the program break continue in order for the station to reach its goal?

Correct solution

1. Program Time During Initial Appeal = $7,000 / $1,200 per minute
2. Program Time During Initial Appeal = 5.83 minutes

3. Amount Collected After Special Appeal = $13,900 - $7,000
4. Amount Collected After Special Appeal = $6,900

5. Program Time After Special Appeal = $6,900 / $3,000 per minute
6. Program Time After Special Appeal = 2.3 minutes

7. Total Program Time Needed = 5.83 minutes + 2.3 minutes
8. Total Program Time Needed = 8 minutes

9. End Time of Program = 6:15 p.m. + 8 minutes
10. End Time of Program = 6:23 p.m.

DRT-08. A machine was set to produce the first 128 electrical components of the 420 needed for a certain production order at a rate of 24 components per hour and the rest of the order at a rate of 40 components per hour. If the machine started filling the order at 7:20 a.m. and ran continuously, at what time, to the nearest minute, was the order completed?

Correct solution

1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours

3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components

5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours

7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
9. Total Production ɪime = 12 hours 38 minutes

10. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
11. Ending Time for Production = 7:58 p.m.

DRT-10. A grocery store sold 317 cans of its brand of tomato sauce yesterday. The first 62 cans were sold at the regular price at the rate of 12 cans per hour. The rest of the cans, which had been marked in error at a lower price, were sold at the rate of 34 cans per hour. If the tomato sauce began selling at 8:00 a.m., at what time, to the nearest minute, was the last can sold?

Correct solution

1. Time to Sell Cans at Regular Price = 62 cans / 12 cans per hour
2. Time to Sell Cans at Regular Price = 5.17 hours

3. Cans Sold at Lower Price = 317 cans - 62 cans
4. Cans Sold at Lower Price = 255 cans

5. Time to Sell Cans at Lower Price = 255 cans / 34 cans per hour
6. Time to Sell Cans at Lower Price = 7.5 hours

7. Total Sales Time = 5.17 hours + 7.5 hours
8. Total Sales Time = 12 hours 40 minutes

9. Ending Time for Sales = 8:00 a.m. + 12 hours 40 minutes
10. Ending Time for Sales = 8:40 p.m.

P-02. In one day, one laborer can pick 0.08 percent of the 900 bushels of fruit that an orchard is expected to yield, which is a "laborer-day" amount picked. How many laborer days would be required to pick all of the fruit that the orchard is expected to yield?

Correct solution A

1. 0.08 Percent = 0.0008

2. Bushels Picked in a Laborer Day = 0.08 percent per day * 900 bushels
3. Bushels Picked in a Laborer Day = 0.72 bushels per laborer day

4. Laborer Days Needed to Pick Targeted Number of Bushels = 900 bushels / 0.72 bushels per labore. day
5. Laborer Days Needed to Pick Targeted Number of Bushels = 1,250 laborer days

Correct solution B

1. Total = 100%

2. 100% of Bushels = 0.08% of bushels per laborer day * X laborer days

3. X Laborer Days = 100% of bushels / 0.08% of bushels per laborer day
4. X Laborer Days = 1,250 laborer days

P-04. A sales representative earns a 4 percent commission on the sale of each encyclopedia set. How many $800 encyclopedia sets must the representative sell to earn a total commission equal to the price of one $800 encyclopedia set?

Correct solution A

1. 4 Percent = 0.04

2. Commission per Set = 0.04 * $800 per set
3. Commission per Set = $32 per set

4. Sets to be Sold to Reach Target Commission = $800 / $32 per set
5. Sets to be Sold to Reach Target Commission = 25 sets

Correct solution B

1. Total = 100%

2. 100% Commission = 4% commission per set * X sets

3. X Sets = 100% commission / 4% commission per set
4. X Sets = 25 sets

P-09. An oil tanker filled to capacity with 10 million barrels of oil is leaking 0.02 percent of its capacity of oil each hour. At this rate, how many hours would it take for all of the tanker's oil to leak out?

Correct solution A

1. 0.02 Percent = 0.0002

2. Amount Leaked per Hour = 0.02 percent per hour * 10 million barrels
3. Amount Leaked per Hour = 0.002 million barrels per hour

4. Time for Total Leakage = 10 million barrels / 0.002 million barrels per hour
5. Time for Total Leakage = 5,000 hours

Correct solution B

1. Total = 100%

2. 100% Leaked = 0.02% leaked per hour * X hours

3. X Hours = 100% leaked / 0.02% leaked per hour
4. X Hours = 5,000 hours

P-12. Geologists estimate that 0.05 percent of a 2-mile-wide reef is eroding each century. At this rate, in how many centuries would the entire reef be eroded?

Correct solution A

1. 0.05 Percent = 0.G005

2. Erosion per Century = 0.05 percent per century * 2 miles
3. Erosion per Century = 0.001 miles per century

4. Time for Complete Erosion = 2 miles / 0.001 miles per century
5. Time for Complete Erosion = 2,000 centuries

Correct solution B

1. Total = 100%

2. 100% Erosion = 0.05% erosion per century * X centuries

3. X Centuries = 100% erosion / 0.05% erosion per century
4. X Centuries = 2,000 centuries

Appendix C

Sample 3 Problems and Canonical Solutions

Problem Statements and Correct Solutions

The following pages contain 12 problems, 4 isomorphs of each of 3 problem types (DRT or Distance=rateXtime, GR or Graduated Rate, and DRT-V or Distance=rateXtime with a variable in the solution).

The lines in each solution are numbered in a way that corresponds to the "Canonical Solutions" used with the "Detailed Error Descriptions."

Lines in the correct solutions are grouped into goals; a blank line separates goals. When there are "A," "B," or "C" versions of a solution, it means that any of the alternative forms is acceptable.

DRT-03. A grocery store sold 317 cans of its brand of tomato sauce yesterday. The first 62 cans were sold at the regular price at the rate of 12 cans per hour. The rest of the cans, which had been marked in error at a lower price, were sold at the rate of 34 cans per hour. If the tomato sauce began selling at 8:00 a.m., at what time, to the nearest minute, was the last can sold?

Correct solution

1. Time to Sell Cans at Regular Price = 62 cans / 12 cans per hour
2. Time to Sell Cans at Regular Price = 5.17 hours

3. Cans Sold at Lower Price = 317 cans - 62 cans
4. Cans Sold at Lower Price = 255 cans

5. Time to Sell Cans at Lower Price = 255 cans / 34 cans per hour
6. Time to Sell Cans at Lower Price = 7.5 hours

7. Total Sales Time = 5.17 hours + 7.5 hours
8. Total Sales Time = 12 hours 40 minutes

9. Ending Time for Sales = 8:00 a.m. + 12 hours 40 minutes
10. Ending Time for Sales = 8:40 p.m.

DRT-05. A machine was set to produce the first 128 electrical components of the 420 needed for a certain production order at a rate of 24 components per hour and the rest of the order at a rate of 40 components per hour. If the machine started filling the order at 7:20 a.m. and ran continuously, at what time, to the nearest minute, was the order completed?

Correct solution

1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours

3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components

5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours

7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 12 hours 38 minutes

9. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
10. Ending Time for Production = 7:58 p.m.

DRT-07. During a break in its program at 6:15 p.m., a TV station began a fund-raising campaign that has a goal of raising $13,900 through viewer solicitation. It raised the first $7,000 at the rate of $1,200 per minute. When a special appeal began, pledges came in at the rate of $3,000 per minute. If funds continue to come in at the same rate as during the special appeal, until what time, to the nearest minute, must the program break continue in order for the station to reach its goal?

Correct solution

1. Program Time During Initial Appeal = $7,000 / $1,200 per minute
2. Program Time During Initial Appeal = 5.83 minutes

3. Amount Collected After Special Appeal = $13,900 - $7,000
4. Amount Collected After Special Appeal = $6,900

5. Program Time After Special Appeal = $6.900 / $3,000 per minute
6. Program Time After Special Appeal = 2.3 minutes

7. Total Program Time Needed = 5.83 minutes + 2.3 minutes
8. Total Program Time Needed = 8 minutes

9. End Time of Program = 6:15 p.m. + 8 minutes
10. End Time of Program = 6:23 p.m.

DRT-12. A fire is burning in a forest of several thousand acres. Firefighters are attempting to confine the fire to one region covering 705 acres, and plan to light a backfire at the moment the fire reaches the edge of that region, after having burned the entire region. The first 459 acres burned at a rate of 90 acres per hour. The wind then diminished, and the fire is now burning at the rate of 36 acres per hour. If the fire started at 10 a.m. and wind conditions do not change, at what time, to the nearest minute, should the backfire be lit?

Correct solution

1. Burning Time1 = 459 acres / 90 acres per hour
2. Burning Time1 = 5.1 hours

3. Acres Amount2 = 705 acres - 459 acres
4. Acres Amount2 = 246 acres

5. Burning Time2 = 246 acres / 36 acres per hour
6. Burning Time2 = 6.83 hours

7. Total Burning Time = 5.1 hours + 6.83 hours
8. Total Burning Time = 11 hours 56 minutes

9. Time Backfire Should Be Lit = 10 a.m. + 11 hours 56 minutes
10. Time Backfire Should Be Lit = 9:56 p.m.

GR-13. A lawyer charges $100 for the first hour of service and $75 for each additional hour. A bill of $625 represents how many hours of the lawyer's service?

____

Correct solution

1. Lawyer's Charge After First Hour = $625 - $100
2. Lawyer's Charge After First Hour = $525

3. Number of Additional Hours = $525 / $75
4. Number of Additional Hours = 7 hours

5. Total Number of Hours = 7 hours + 1 hour
6. Total Number of Hours = 8 hours

GR-15. An entertainer's contract specifies payment of $25,000 for the first performance and $18,000 for each additional performance. If the contract specifies a total payment of $277,000 for the entertainer's performances, how many performances are required?

Correct solution

1. Contract Charge After First Performance = $277,000 - $25,000
2. Contract Charge After First Performance = $252,000

3. Number of Additional Performances = $252,000 / $18,000
4. Number of Additional Performances = 14

5. Total Number of Performances = 14 performances + 1 performance
6. Total Number of Performances = 15 performances

GR-18. A taxi charges $2.00 for the first mile and $1.40 for each additional mile. If the total charge is $18.80 for a certain trip, how many miles long is the trip?

Correct solution

1. Charge After First Mile of Trip = $18.80 - $2.00
2. Charge After First Mile of Trip = $16.80

3. Miles After First Mile of Trip = $16.80 / $1.40
4. Miles After First Mile of Trip = 12 miles

5. Total Miles for Trip = 1 mile + 12 miles
6. Total Miles for Trip = 13 miles

GR-20. The charge to rent a boat is $20 for the first hour and $12 for each additional hour. If the total charge to rent a boat was $116, for how many hours was the boat rented?

Correct solution

1. Rental Charge After First Hour = $116 - $20
2. Rental Charge After First Hour = $96

3. Number of Rental Hours After First Hour = $96 / $12
4. Number of Rental Hours After First Hour = 8 hours

5. Total Number of Hours = 8 hours + 1 hour
6. Total Number of Hours = 9 hours

DRT-V-14. A car is traveling at an average speed of 80 kilometers per hour. On the average, how many seconds does it take the car to travel K kilometers?

Correct solution A

1. Kilometers Traveled per Second = 80 kilometers per hour / 3,600 seconds per hour
2. Kilometers Traveled per Second = 1/45 kilometers per second

3. Seconds to Travel K Kilometers = K kilometers / 1/45 kilometer per second
4. Seconds to Travel K. Kilometers = 45K seconds

Correct solution B

1. Hours to Travel K Kilometers = K kilometers / 80 kilometers per hour
2. Hours to Travel K Kilometers = K/80 hours

3. Seconds to Travel K Kilometers = K/80 hours * 3,600 seconds per hour
4. Seconds to Travel K Kilometers = 45K seconds

Correct solution C

1. Seconds per Kilometer = 1/80 hour per kilometer * 3,600 seconds per hour
2. Seconds per Kilometer = 45 seconds per kilometer

3. Seconds to Travel K Kilometers = 45 seconds per kilometer * K kilometers
4. Seconds to Travel K Kilometers = 45K seconds

DRT-V-16. A local phone system processes an average of 12,000 calls each hour. On the average, how many seconds would it take the phone system to process K calls?

Correct solution A

1. Calls Processed per Second = 12,000 calls processed per hour / 3,600 seconds per hour
2. Calls Processed per Second = 3 1/3 calls processed per second

3. Seconds to Process K Calls = K calls / 3 1/3 calls processed per second
4. Seconds to Process K Calls = 3K/10 seconds

Correct solution B

1. Hours to Process K Calls = K calls / 12,000 calls per hour
2. Hours to Process K Calls = K/12,000 hours

3. Seconds to Process K Calls = K/12,000 hours * 3,600 seconds per hour
4. Seconds to Process K Calls = 3K/10 seconds

Correct solution C

1. Seconds per Call = 1/12,000 hour per call * 3,600 seconds per hour
2. Seconds per Call = 3/10 seconds per call

3. Seconds to Process K Calls = 3/10 seconds per call * K calls
4. Seconds to Process K Calls = 3K/10 seconds

DRT-V-17. A taxicab driver earns on average $144 for each eight-hour shift. On average, how many hours does it ta e the driver to earn C dollars?

Correct solution A

1. Dollars Earned per Hour = $144 per shift / 8 hours per shift
2. Dollars Earned per Hour = $18.00 per hour

3. Hours to Earn C Dollars = C dollars / $18.00 per hour
4. Hours to Earn C Dollars = C /18 hours

Correct solution B

1. Number of Shifts to Earn C Dollars = C dollars / $144 per shift
2. Number of Shifts to Earn C Dollars = C/144 shifts

3. Hours to Earn C Dollars = C/144 shifts * 8 hours per shift
4. Hours to Earn C Dollars = C/18 hours

Correct solution C

1. Hours per Dollar Earned = 1/144 shifts per dollar * 8 hours per shift
2. Hours per Dollar Earned = 1/18 hours per dollar

3. Hours to Earn C Dollars = 1/18 hours per dollar * C dollars
4. Hours to Earn C Dollars = C/18 hours

DRT-V-19. Each hour, an average of 8.000 letters are sorted by a postal machine. At that rate, how many seconds does it take, on average, for the machine to sort L letters?

Correct solution A

1. Letters Sorted per Second = 8.000 letters per hour / 3,600 seconds per hour
2. Letters Sorted per Second = 20/9 letters per second

3. Seconds to Sort L Letters = L letters / 20/9 letters per second
4. Seconds to Sort L Letters = 9L/20 seconds

Correct solution B

1. Hours to Sort L Letters = L letters / 8,000 letters per hour
2. Hours to Sort L Letters = L/8,000 hours

3. Seconds to Sort L Letters = L/8,000 hours * 3,600 seconds per hour
4. Seconds to Sort L Letters = 9L/20 seconds

Correct solution C

1. Seconds per Letter = 1/8,000 hour per letter * 3,600 seconds per hour
2. Seconds per Letter = 9/20 seconds per letter

3. Seconds to Sort L Letters = 9/20 seconds per letter * L letters
4. Seconds to Sort L Letters = 9L/20 seconds

Appendix D

Sample 2 Judges' Instructions

Diagnostic Analysis Instructions

I. Task Description

A. We would like you to read a set of student solutions to 12 algebra word problems and to identify any errors in them by (1) circling each error and (2) associating it with a code. Some solutions may be error free, others may contain multiple errors. In some cases, a single statement or value may reflect multiple combined errors. In all cases, you should identify each of the errors involved. If a portion of a solution can be interpreted to suggest alternative error explanations, pick the error that you think is most probably the correct explanation.

II. Problem Description

A. The problems you will be reading fall into 3 problem types: Distance=RatexTime (DRT), PERCENT, and WORK. Each problem type requires a specif c number of goals to be achieved in order to solve the problem: The DRT problems have 5 goals, the I RECENT problems 3 goals, and the WORK problems 2 goals.

B. For each problem type (DRT, PRECENT,WORK), there are 4 isomorphic problems. These isomorphs differ in specific content and the "format" or amount of additional information provided, but they preserve the same solution structure. The same goals and procedures (with new labels and values) can therefore apply to each of the 4 ismorophic problems of a single problem type. You will notice the differences in the problem "formats," but those differences need not play a role in your task. Each solution should be evaluated in the same way, regardless of the format of problem presentation.

III. Lists of Materials

A. To facilitate your task, we are providing you with the following materials:

1. A "Problem Statement" for each of the 12 problems presented to the examinees, together with a correct solution. The correct solution is broken down into separate goals that need to be accomplished to reach a solution as well as an example of a specific set of correct procedures.

2. A "Bug Summary" for each problem type that provides a quick reference to the bug codes you are to use

3. A set of "Detailed Error Descriptions and Examples" that provides additional explanation for the "Bug Summaries"

4. A set of "Bug-Not-Listed" forms for recording any errors that do not fit the bug types listed in the bug summary

5. Examples of Student Solutions for which errors have already been identified

IV. Specific Procedure

A. For each problem, locate each uncorrected error and circle it. If the error is one of those listed on the "Bug Summary," write the three-digit code number for that bug next to the error.

B. Missing goal bugs should be used only when there is no reasonable attempt to achieve the goal. A goal need not be represented explicitly, if it can be determined from other parts of the problem that it was satisfied. For example, if a student writes "$800 * 0.04 = $32.00" for a problem that requires 4% of $800, we can assume that the student successfully completed the goal of converting 4% to 0.04 even though the conversion is not written down. If a missing goal error is used (900s), no other error can be assigned to that goal.

C. In those cases for which there is no appropriate error in the "Bug Summary" list, write the code 999 next to the error. Then write the subject number, problem number, and a description of the error on the "Bugs Not Listed" form.

D. If a solution has no errors, simply write OK and circle it on the solution. In general, you should give the student the benefit of a doubt in evaluating the solution. If an answer is correct, it should be considered to be error-free unless there is clear evidence to the contrary.

E. Use the most specific bug available.

1. Use Specific Plan Bugs before General Plan Bugs.

   a) For example, if a student writes "128/24 = 5 hours 33 minutes," DECIMAL PORTION TREATED AS TIME (#302) would always be used before UNITS MATCH, TENTHS DO NOT (#111). Even though (#111) is a correct description, you should use (#302) since it is more specific.

2. Use most specific math bugs first.

   a) For example, if a student writes "5.3 instead of 5.33," TENTHS MATCH, HUNDREDTHS DO NOT (#110) would always be used before UNITS MATCH, TENTHS DO NOT (#111).

F. Distinguish single from multiple errors.

1. The following example has a single error (#302) in which a student has converted a decimal representation to time (in this case, minutes):

   Time to Sell Cans at Regular Price = 5.17 hours
   Time to Sell Cans at Regular Price = 5 hours 17 minutes

2. The next example, in contrast, has two separable errors, one rounding 5.1666 to 5.16 instead of 5.17 (#110), and the second in confusing decimal and time representations as in the previous example (#302):

   Time to Sell Cans at Regular Price = 62/12 hours
   Time to Sell Cans at Regular Price = 5 hours 16 minutes

G. Count each error only once.

1. If a student has made a computational error, use the student's incorrect value to evaluate the remainder of the solution.

H. Precision is required only to 2 decimal places for any part of the solution.

Appendix E

Sample 2 Bug Classification Scheme and Detailed Error Descriptions with Examples

DRT Bug Summary

Math Bugs

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33
(111) Units Match, Tenths Do Not: 7.2 instead of 7.3
(113) Decimal Shift: 53.3 instead of 5.33
(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours
(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 6.2

Specific Plan Bugs

DRT Time Bugs

(301) Division Remainder Treated as Time: 128 Components / 24 Components per Hour= 5 hours 8 minutes
(302) Decimal Portion Treated as Time: 128 Components / 24 Components per Hour = 5 hours 33 minutes
(304) Time Treated as a Decimal: 128 Components / 24Components per Hour = 5.20 hours
(305) Shift A.M. to P.M. or P.M. to A.M.: Ending Time = 7:20 a.m. + 12 hours 38 minutes = 7:58 a.m.
(306) Not Exact Match, but Within 1 Minute: Production Time1 = 5.34 hours instead of 5.33 hours
(307) Use Decimal for Colon in Time: 7.58 instead of 7:58

Other Specific Plan Bugs

(401) Wrong Rate Value in a Structurally Correct Plan: Time1 = 128 components / 44 components per hour
(402) Wrong Quantity Value in a Structurally Correct Plan: Quantity2 = 450 components - 128 components
(404) Rate2 Used for Rate1: Production Time1 = 128 components / 40 components per hour
(405) Rate1 Used for Rate2: Production Time2 = 292 components / 24 components per hour
(406) Total Quantity Used for Partial Quantity: Production Time2 = 420 components / 40 components per hour
(407) Use Partial Instead of Total Elapsed Time: Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
(411) Unknown Value Used for Time1: Total Production Time = 5.33 hours + 6.3 hours
(412) Unknown Value Used for Time2: Total Production Time = 4 hours + 7.3 hours
(413) Unknown Value Used for Total Time: Ending Time for Production = 7:20 a.m. + 10 hours
(414) Average Rate Not Weighted by Time: Average Rate = (24 components per hour + 40 components per hour)/2

General Plan Bugs

(701) Expression Not Reduced: Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
(702) Final Goal Not Reduced: Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
(703) Multiplication Used Where Division Required: Time1 = 128 components * 24 components per hour
(704) Division Used Where Multiplication Required: Commission per Set = 0.04 / $800 per set
(705) Addition Used Where Subtraction Required: Quantity2 = 420 components + 128 components
(710) Subtraction Used Where Addition Required: Total Production Time = 7.3 hours - 5.33 hours
(707) Correct Plan Structure, Multiple Unknown Values: Total Production Time = 6.6 hours + 6.3 hours
(708) Correct Plan Structure, Single Unknown Value: Time1 = 128 components / 45 components per hour
(709) Label Right, Value Close; No Specific Explanation: Quantity2 = 280 components

Missing DRT Goals

(911)	Missing First Goal:	Time1 = Quantity1 / Rate1
(912)	Missing Second Goal:	Quantity2 = Total Quantity - Quantity1
(913)	Missing Third Goal:	Time2 = Quantity2 / Rate2
(914)	Missing Fourth Goal:	Total Time = Time1 + Time2
(915)	Missing Fifth Goal:	Finish Time = Start Time + Total Time

PERCENT Bug Summary

Math Bugs

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33
(111) Units Match, Tenths Do Not: 7.2 instead of 7.3
(113) Decimal Shift: 53.3 instead of 5.33
(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours
(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 6.2

Specific Plan Bugs

Percent Bugs

(501) Treat Percent as a Decimal: Commission per Set = 4 * $800 (4% as 4.0)
(502) Treat Decimal as Percent: 0.04 * $800 = $0.32 (0.04 as 0.0004)
(503) Mix Percent and Decimal Values: 100% Commission / 0.04 Commission per Set = 2,500 sets
(504) Unit-Return as Unit-Value/Rate instead of Unit-Value * Rate: Commission per Set = $800 / 0.04
(505) Percent as Decimal Denominator of Fraction: $800 * 4% (= $800 * 1/4) = $800 * .25

General Plan Bugs

(701) Expression Not Reduced: Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
(702) Final Goal Not Reduced: Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
(703) Multiplication Used Where Division Required: Time1 = 128 components * 24 components per hour
(704) Division Used Where Multiplication Required: Commission per Set = 0.04 / $800 per set
(705) Addition Used Where Subtraction Required: Quantity2 = 420 components + 128 components
(710) Subtraction Used Where Addition Required: Total Production Time = 7.3 hours - 5.33 hours
(707) Correct Plan Structure, Multiple Unknown Values: Total Production Time = 6.6 hours + 6.3 hours
(708) Correct Plan Structure, Single Unknown Value: Time1 = 128 components / 45 components per hour
(709) Label Right, Value Close; No Specific Explanation: Quantity2 = 280 components

Missing PERCENT Goals

(921)	Missing First Goal:	Decimal = .01 * percent
(922)	Missing Second Goal:	Unit-Return = Rate Per Unit * Unit-Value
(923)	Missing Third Goal:	Time = Unit-Value / Unit-Return

WORK Bug Summary

Math Bugs

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33
(111) Units Match, Tenths Do Not: 7.2 instead of 7.3
(113) Decimal Shift: 53.3 instead of 5.33
(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours
(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 5.2

Specific Plan Bugs

Work Bugs

(603) Subtract Items in Wrong Order for Net Rate: Nondefective Bottles per Minute = 3 - 35 bottles per minute
(604) Increase/Loss Ratio for Net Rate: Nondefective Bottles per Minute = 35 bottles per minutes / 3 bottles per minute
(605) Increase/Loss * Increase for Net Rate: Nondefective Bottles per Minute = (35 / 3) * 35 bottles per minute
(606) Multiply for Subtract in Net Rate: Nondefective Bottles per Minute = 35 bottles per minute * 3 bottles per minute
(607) Increase Rate for Net Rate: Time for Desired Bottles = 8,000 bottles / 35 bottles per minute
(608) Loss Rate for Net Rate: Time for Desired Bottles = 8,000 bottles / 3 bottles per minute

General Plan Bugs

(701) Expression Not Reduced: Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
(702) Final Goal Not Reduced: Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
(703) Multiplication Used Where Division Required: Time1 = 128 components * 24 components per hour
(704) Division Used Where Multiplication Required: Commission per Set = 0.04 / $800 per set
(705) Addition Used Where Subtraction Required: Quantity2 = 420 components + 128 components
(710) Subtraction Used Where Addition Required: Total Production Time = 7.3 hours - 5.33 hours
(707) Correct Plan Structure, Multiple Unknown Values: Total Production Time = 6.6 hours + 6.3 hours
(708) Correct Plan Structure, Single Unknown Value: Time1 = 128 components / 45 components per hour
(709) Label Right, Value Close; No Specific Explanation: Quantity2 = 280 components

Missing WORK Goals

(931) Missing First Goal:	Net Rate = Increase Rate - Loss Rate
(932) Missing Second Goal:	Time = Quantity / Net Rate

Detailed Error Descriptions and Examples

This section provides more detailed descriptions of the errors listed on the "Bug Summary" pages.

Three canonical solutions, identical to those from the "Problem Statements and Correct Solutions" section, are repeated here for convenience. Examples of errors are shown as deviations from these canonical solutions. The solution from which the example deviates is indicated by DRT for distance problems, % for percent problems, and WORK for work problems. In most cases only the relevant modified lines are given, and they are numbered to match the canonical solution. For a number of examples, additional lines are provided for context. Lines on which the error occurs are marked by an asterisk. In a few cases, intermediate steps that are not usually shown in the solution are presented in parentheses to clarify the example.

Canonical Solutions

(Roman Numerals Indicate Goals)

DRT (Five-Goal Problems)

I. 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours

II. 3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components

III. 5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours

IV. 7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 12 hours 38 minutes

V. 9. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
10. Ending Time for Production = 7:58 p.m.

Percent (%) (Three-Goal Problems)

Solution A:

I. 1. 4 Percent = 0.04

II. 2. Commission per Set = 0.04 * $800 per set
3. Commission per Set = $32 per set

III. 4. Sets to be Sold to Reach Target Commission = $800 / $32 per set
5. Sets to be Sold to Reach Target Commission = 25 sets

Solution B:

I. 1. Total = 100%

II. 2. 100% Commission = 4% commission per set * X sets

III. 3. X Sets = 100% commission / 4% commission per set
4. X Sets = 25 sets

Percent conversion and annual dividends are not required in this approach. Since the solution is considered more elegant and achieves the same objective as strategy A, it is given the same number of total goals. The first goal is rarely stated explicitly, but is implicit in the second one.

Work (Two-Goal Problems)

I. 1. Nondefective Bottles per Minute = 35 bottles per minute - 3 bottles per minute
2. Nondefective Bottles per Minute = 32 bottles per minute

II. 3. Time for Desired Nondefective Bottles = 8,000 bottles / 32 bottles per minute
4. Time for Desired Nondefective Bottles = 250 minutes

Detailed Error Descriptions

Math Bugs

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33

DRT 2. Production Time1 = 5.30 hours

(111) Units Match, Tenths Do Not: 7.2 instead of 7.3

Errors in the hundredths only (bug #110) are checked before errors in tenths only (this bug #111).

DRT 6. Production Time2 = 7.2 hours

(113) Decimal Shift: 53.3 instead of 5.33

The obtained value has a shifted decimal with respect to the expected value.

DRT 2. Production Time1 = 53.3 hours

(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours

(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 6.2

This error is used for any equation in which a complex expression is set equal to an incorrect value that cannot be ascribed to a "Specific Plan Bug" or to a more specific "Math Bug."

Specific Plan Bugs: DRT Time Bugs

(301) Division Remainder Treated as Time:

DP.T 1. Production Time1 = 128 components / 24 components per hour
(Production Time1 = 5 remainder 8 hours)
*2. Production Time1 = 5 hours 8 minutes

(302) Decimal Portion Treated as Time:

DRT 1. Production Time1 = 128 components / 24 components per hour
(Production Time1 = 5.33 hours)
*2. Production Time1 = 5 hours 33 minutes

(304) Time Treated as a Decimal:

DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5 hours 20 minutes
*. Production Time1 = 5.20 hours

(305) Shift A.M. to P.M. or P.M. to A.M.:

This bug is sometimes triggered by student failure to indicate either a.m. or p.m., in which case a.m. is assumed.

DRT 11. Ending Time for Production = 7:58 a.m.

(306) Not Exact Match, but Within 1 Minute:

DRT . 2. Production Time1 = 5.34 hours

(307) Use Decimal for Colon in Time:

A clock time is represented with a decimal in place of a colon.

DRT 10. Ending Time for Production = 7.58

Specific Plan Bugs: Other DRT Bugs

(401) Wrong Rate Value in a Structurally Correct Plan:

DRT 1. Production Time1 = 128 components / 44 components per hour

(402) Wrong Quantity Value in a Structurally Correct Plan:

DRT 3. Component Quantity2 = 450 components - 128 components

(404) Rate2 Used for Rate1:

DRT 1. Production Time1 = 128 components / 40 components per hour

(405) Rate1 Used for Rate2:

DRT 5. Production Time2 = 292 components / 24 components per hour

(406) Total Quantity Used for Partial Quantity:

DRT 5. Production Time2 = 420 components / 40 components per hour

(407) Use Partial Instead of Total Elapsed Time:

The ending time uses only one of the elapsed times instead of the total elapsed time.

DRT 10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes

(411) Unknown Value Used for Time1:

The structure of the solution suggests that Time1 is being used, but the source of the value used is unknown.

DRT *7. Total Production Time = 4 hours + 7.3 hours
8. Total Production Time = 4 hours + 7 hours 18 minutes
9. Total Production Time = 11 hours 18 minutes
10. Ending Time for Production = 7:20 a.m. + 11 hours 18 minutes
11. Ending Time for Production = 6:38 p.m.

(412) Unknown Value used for Time2:

The structure of the solution suggests that Time2 is being used, but the source of the value used is unknown.

DRT *7. Total Production Time = 5.33 hours + 6 hours
8. Total Production Time = 5 hours 20 minutes + 6 hours
9. Total Production Time = 11 hours 20 minutes
10. Ending Time for Production = 7:20 a.m. + 11 hours 20 minutes
11. Ending Time for Production = 6:40 p.m.

(413) Unknown Value used for Total Time:

DRT 10. Ending Time for Production = 7:20 a.m. + 10 hours
11. Ending Time for Production = 7:58 p.m.

(414) Average Rate Not Weighted by Time:

An average rate is computed improperly since it is not weighted according to the different times. This example assumes a nonstandard approach to the problem.

DRT Average Rate = (24 components per hour + 40 components per hour)/2
Average Rate = 32 components per hour
Average Time = 420 components / 32 components per hour
Average Time = 13.12 hours

<u>Specific Plan Bugs: Percent Bugs</u>

(501) Treat Percent as a Decimal:

E.g., 4% treated as 4.0

%A 1. absent
2. Commission per Set = 4% * $800 per set
3. Commission per Set = $3,200 per set
4. Sets to be Sold to Reach Target Commission = $800 / $3,200 per set
5. Sets to be Sold to Reach Target Commission = 0.25 sets

(502) Treat Decimal as Percent:

E.g., .04 as if it were .04% or .0004.

%A 2. Commission per Set = 0.04 * $800
3. Commission per Set = $0.32

(503) Mix Percent and Decimal Values:

%B 2. 100% Commission = 0.04 commission per set * X sets
3. X Sets = 100% commission / 0.04 commission per set
4. X Sets = 2,500 sets

(504) Unit-Return as Unit-Value/Rate instead of Unit-Value * Rate:

The unit-return is calculated as a division of unit-value by rate instead of a multiplication of unit-value times the rate.

%A 2. Commission per Set = $800 per set / 0.04

(505) Percent as Decimal Denominator of Fraction:

The percent value is treated as the denominator of a fraction (e.g. 4% is interpreted as 1/4).

%A 2. Commission per Set = 4% * $800 per set
3. Commission per Set = $200

Specific Plan Bugs: Work Bugs

(603) Subtract Items in Wrong Order for Net Rate:

E.g., Defective Bottles minus Nondefective Bottles

WORK 1. Nondefective Bottles per Minute = 3 bottles per minute - 35 bottles per minute

(604) Increase/Loss Ratio for Net Rate:

A ratio is used instead of a subtraction to determine the net rate.

WORK 1. Nondefective Bottles per Minute = 35 bottles per minute / 3 bottles per minute

(605) Increase/Loss * Increase for Net Rate:

A specific increase and loss formulation is used instead of a subtraction for a net rate.

WORK 1. Nondefective Bottles per Minute = (35 / 3) * 35 bottles per minute

(606) Multiply for Subtract in Net Rate:

WORK 1. Nondefective Bottles per Minute = 3 bottles per minute * 35 bottles per minute

(607) Increase-Rate for Net Rate:

WORK 3. Time for Desired Nondefective Bottles = 8,000 bottles / 35 bottles per minute

(608) Loss-Rate for Net Rate:

WORK 3. Time for Desired Nondefective Bottles = 8.000 bottles / 3 bottles per minute

General Plan Bugs

(701) Expression Not Reduced:

An expression is not sufficiently reduced. This bug is reported only if the nonreduced value is not resolved later in the solution.

DRT 8. Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
No subsequent lines in solution

(702) Final Goal Not Reduced:

The "final" answer to the problem is not reduced. It is like the no-reduction bug, but applies to the final goal.

DRT 10. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
No subsequent lines in solution

(703) Multiplication Used Where Division Required:

DRT 1. Production Time1 = 128 components * 24 components per hour

(704) Division Used Where Multiplication Required:

%A 2. Commission per Set = 0.04 / $800 per set

(705) Addition Used Where Subtraction Required:

DRT 3. Component Quantity2 = 420 components + 128 components

(707) Correct Plan Structure, Unknown Values:

The student uses a value of unknown origin, but it fits into the problem solution structure for subsequent goals. Unlike bug #708, this bug is not restricted to a single value. The correctness of the plan is usually recognized because of the appropriate use in subsequent goals.

DRT *7. Total Production Time = 6.6 hours + 6.3 hours
8. Total Production Time = 6 hours 36 minutes + 6 hours 18 minutes
9. Total Production Time = 12 hours 54 minutes
10. Ending Time for Production = 7:20 a.m. + 12 hours 54 minutes
11. Ending Time for Production = 8:14 p.m.

(708) Correct Plan Structure, Single Unknown Value:

A plan has a single unexplained value within the correct structure. This is in contrast to bug #707, which includes errors in multiple values. This bug is used primarily in INTEREST and WORK problems. DRT problems use more specific bugs, 401 and 402.

WORK 3. Time for Desired Nondefective Bottles = 8,000 bottles / 65 bottles per minute

(7C9) Label Right, Value Close; No Specific Explanation:

This error is used when the label is the only basis for identifying what the student is doing. It indicates that the combination of other plans and bugs could not adequately explain the value. At the same time, the value assigned to the bug must be within a reasonable range of the expected value, in this case within 1% of the larger of the obtained and expected values. Notice that this is not considered a math bug (cf #112) because there is no explicit statement of how the student derived the value.

DRT 3. absent
4. Component Quantity2 = 280 components

(710) Subtraction Used Where Addition Required:

As with other bugs of this general form, this bug should be identified only when it is seen as a transformation of an otherwise "correct" plan. If the values and operator are jointly wrong, then there is a missing goal.

DRT 9. Total Production Time = 7.3 hours - 5.33 hours

Missing Goals

These bugs indicate that the stated goal is missing. No goal can have any other associated bug if it is "missing."

DRT Missing Goals

(911) Missing First Goal, Time1 = Quantity1 / Rate1
DRT 3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
10. Ending Time for Production = 7:20 a.m. + 7 hours 18 minutes
11. Ending Time for Production = 2:38 p.m.

(912) Missing Second Goal, Quantity = Total Quantity - Quantity1
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
5. Production Time2 = 128 components / 40 components per hour
6. Production Time2 = 3.2 hours
7. Total Production Time = 5.33 hours + 3.2 hours
8. Total Production Time = 5 hours 20 minutes + 3 hours 12 minutes
9. Total Production Time = 8 hours 32 minutes
10. Ending Time for Production = 7:20 a.m. + 8 hours 32 minutes
11. Ending Time for Production = 3:52 p.m.

(913) Missing Third Goal, Time2 = Quantity2 / Rate2
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
11. Ending Time for Production = 12:40 p.m.

(914) Missing Fourth Goal, Total Time = Time1 + Time2
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
11. Ending Time for Production = 12:40 p.m.

(915) Missing Fifth Goal, Finish Time = Start Time + Total Time
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
9. Total Production Time = 12 hours 38 minutes

Percent Missing Goals

(921) Missing First Goal, Decimal = .01 * percent

For the current set of problems, this bug can be the same as #501. Since "Treat Percent as a Decimal" (#501) makes a more specific claim, it should be used whenever there is specific evidence of misconversion. The "Missing First Goal" bug will still be needed in those cases in which parts of the problem are left out but it is not clear why.

%A 2. Commission per Set = 4% * $800 per set
3. Commission per Set = $3,200 per set
4. Sets to be Sold to Reach Target Commission = $800 / $3,200 per set
5. Sets to be Sold to Reach Target Commission = 0.25 sets

(922) Missing Second Goal, Unit-Return = Rate Per Unit * Unit-Value

%A 4. Sets to be Sold to Reach Target Commission = $800 / 4
5. Sets to be Sold to Reach Target Commission = 200 sets

(923) Missing Third Goal, Time = Unit-Value / Unit-Return

%A 1. 4 Percent = 0.04
2. Commission per Set = 0.04 * $800
3. Commission per Set = $32

Work Missing Goals

(931) Missing First Goal, Net Rate = Increase Rate - Loss Rate

WORK 3. Time for Desired Nondefective Bottles = 8,000 bottles / 35 bottles per minute
4. Time for Desired Nondefective Bottles = 228.57 minutes

(932) Missing Second Goal, Time = Quantity / Net Rate

WORK 1. Nondefective Bottles per Minute = 35 bottles per minute - 3 bottles per minute
2. Nondefective Bottles per Minute = 32 bottles per minute

Appendix F

Sample 3 Bug Classification Scheme and Detailed Error Descriptions with Examples

DRT Bug Summary

Math Bugs

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33
(111) Units Match, Tenths Do Not: 7.2 instead of 7.3
(113) Decimal Shift: 53.3 instead of 5.33
(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours
(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 6.2
(116) Decimal Treated as Fraction: 12.6 as 12 1/6 hours

Specific Plan Bugs

DRT Time Bugs

(301) Division Remainder Treated as Time: 128 Components / 24 Components per Hour= 5 hours 8 minutes
(302) Decimal Portion Treated as Time: 128 Components / 24 Components per Hour = 5 hours 33 minutes
(304) Time Treated as a Decimal: 128 Components / 24 Components per Hour = 5.20 hours
(305) Shift A.M. to P.M. or P.M. to A.M.: Ending Time = 7:20 a.m. + 12 hours 38 minutes = 7:58 a.m.
(306) Not Exact Match, but Within 1 Minute: Production Time1 = 5.34 hours instead of 5.33 hours
(307) Use Decimal for Colon in Time: 7.58 instead of 7:58
(362) T = R/D used for T = D/R

Other Specific Plan Bugs

(401) Wrong Rate Value in a Structurally Correct Plan: Time1 = 128 components / 44 components per hour
(402) Wrong Quantity Value in a Structurally Correct Plan: Quantity2 = 450 components - 128 components
(404) Rate2 Used for Rate1: Production Time1 = 128 components / 40 components per hour
(405) Rate1 Used for Rate2: Production Time2 = 292 components / 24 components per hour
(406) Total Quantity Used for Partial Quantity: Production Time2 = 420 components / 40 components per hour
(407) Use Partial Instead of Total Elapsed Time: Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
(411) Unknown Value Used for Time1: Total Production Time = 5.33 hours + 6.3 hours
(412) Unknown Value Used for Time2: Total Production Time = 4 hours + 7.3 hours
(413) Unknown Value Used for Total Time: Ending Time for Production = 7:20 a.m. + 10 hours
(414) Average Rate Not Weighted by Time: Average Rate = (24 components per hour + 40 components per hour)/2
(415) Minute Tota' Treated as Increment: 7:20 A.M. + 12 Hours 38 Minutes = 7:20 + 12 hours + 58 minutes = 8:18 p.m.
(416) Hours Treated as Minutes: 7:20 A.M. + 12.63 (Hours) = 7:33 a.m.
(417) Minutes Treated as Hours: 6:15 P.M. + 5.83 (Minutes) = 12:05 a.m.

General Plan Bugs

(701) Expression Not Reduced: Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
(702) Final Goal Not Reduced: Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
(703) Multiplication Used Where Division Required: Time1 = 128 components * 24 components per hour
(704) Division Used Where Multiplication Required: Commission per Set = 0.04 / $800 per set
(705) Addition Used Where Subtraction Required: Quantity2 = 420 components + 128 components
(710) Subtraction Used Where Addition Required: Total Production Time = 7.3 hours - 5.33 hours
(707) Correct Plan Structure, Multiple Unknown Values: Total Production Time = 6.6 hours + 6.3 hours
(708) Correct Plan Structure, Single Unknown Value: Time1 = 128 components / 45 components per hour
(709) Label Right, Value Close; No Specific Explanation: Quantity2 = 280 components

Missing DRT Goals

(911)	Missing First Goal:	Time1 = Quantity1 / Rate1
(912)	Missing Second Goal:	Quantity2 = Total Quantity - Quantity1
(913)	Missing Third Goal:	Time2 = Quantity2 / Rate2
(914)	Missing Fourth Goal:	Total Time = Time1 + Time2
(915)	Missing Fifth Goal:	Finish Time = Start Time + Total Time

GRADUATED RATE Bug Summary

Math Bugs

(109) Division Operator with Multiplication Operation: 96/12 = 1152

(110) Tenths Match, Hundredths Do Not: 1.45 instead of 1.40

(111) Units Match, Tenths Do Not: 8.2 instead of 8.0

(113) Decimal Shift: 960 instead of 96

(114) Remainder of Division Treated as a Decimal: 116 / 12 = 9 remainder 8 = 9.8

(115) Computation Error, Not Identified by Other Math or Plan Errors: 96 / 12 = 9

Specific Plan Bugs

Graduated Rate Bugs

(802) Use First Rate as Only Rate: Time = $625 / $100 per hour = 6.25 hours

(804) Rate Difference as Average Rate: Rate = $100 per hour - $75 per hour = $25 per hour

(805) Rate Average as Hourly Cost: Rate = ($100 + $75) / 2 = $87.50 per hour

Other Specific Plan Bugs

(401) Wrong Rate Value in a Structurally Correct Plan: Time = $525/ $80 per hour

(402) Wrong Quantity Value in a Structurally Correct Plan: Charge = $625 - $200

(405) Rate1 Used for Rate2: Time = $525/ $100 per hour

(406) Total Quantity Used for Partial Quantity: Time = $625 / $75 per hour = 8.33 hours

General Plan Bugs

(701) Expression Not Reduced: Time = $525 / $75 per hour

(702) Final Goal Not Reduced: Total Hours = 7 + 1

(703) Multiplication Used Where Division Required: Time = $525 * $100 per hour

(705) Addition Used Where Subtraction Required: Charges after 1st Hour = $625 + $100

(710) Subtraction Used Where Addition Required: Total Hours = 7 - 1

(708) Correct Plan Structure, Single Unknown Value: Time = $725 / $75 per hour

(709) Label Right, Value Close; No Specific Explanation: Total Hours = 9 hours

Missing DRT Goals

(941) Missing First Goal (Added Amount)	Amount after 1st Hour = Total - 1st hour (Charge after 1st hour = $625 - $100)
(942) Missing Second Goal (Added Hours)	Added Hours = Amount after 1st Hour / Rate after 1st Hour (Added Hours = $525 / $75)
(943) Missing Third Goal (Total Hours)	Total Hours = Added Hours + 1st Hour (Total Hours = 7 hours + 1 hour)

DISTANCE=RATExTIME - VARIABLE Bug Summary

Math Bugs

(109) Division Operator with Multiplication Operation: K/1/45 = K/45 instead of 45K
(110) Tenths Match, Hundredths Do Not: 2.2 instead of 2.22
(111) Units Match, Tenths Do Not: 2.3 instead of 2.22
(113) Decimal Shift: 22.2 instead of 2.22
(114) Remainder of Division Treated as a Decimal: 80/45 = 1 remainder 35 = 1.35
(115) Computation Error, Not Identified by Other Math or Plan Errors: 80/3,600 = 1/40

Specific Plan Bugs

DRT Time Bugs

(301) Division Remainder Treated as Time: 128 Components / 24 Components per Hour= 5 hours 8 minutes
(302) Decimal Portion Treated as Time: 128 Components / 24 Components per Hour = 5 hours 33 minutes
(304) Time Treated as a Decimal: 128 Components / 24 Components per Hour = 5.20 hours
(306) Not Exact Match, but Within 1 Minute: Production Time1 = 5.34 hours instead of 5.33 hours
(351) Convert to Minutes instead of Seconds: 80/60 instead of 80/3,600
(352) Invert Time Conversion: 80 * 3,600 instead of 80/3,600
(361) T=RD for T=D/R: Time = K kilometers * 1/45 kph instead of Time = K/ 1/45
(362) T=R/D for T=D/R: Time = 1/45 kph / K kilometers instead of Time = K/ 1/45
(363) T= (1/R) / D: (1/3.33 Calls per Second) / K = .3 seconds / K

Other Specific Plan Bugs

(401) Wrong Rate Value in a Structurally Correct Plan: Time = K kilometers / 3
(402) Wrong Quantity Value in a Structurally Correct Plan: Kilometers per Second = 100 / 3,600
(421) Treat Constant as 1: Travel Time for K kilometers = 1 kilometer / 1/45 kilometers per second

General Plan Bugs

(701) Expression Not Reduced: Kilometers = 80 / 3,600
(702) Final Goal Not Reduced: Time = K / 1/45
(703) Multiplication Used Where Division Required:
(704) Division Used Where Multiplication Required: Time = K/80 / 3,600
(707) Correct Plan Structure, Multiple Unknown Values:
(708) Correct Plan Structure, Single Unknown Value: Time = 5K kilometers / 0.022
(709) Label Right, Value Close; No Specific Explanation: Time = 50 seconds
(711) Set Correct Response to Wrong Value: C/18 = 144

Missing DRT Goals

(951) Missing First "A" Goal (Rate Convert): Rate_per_Sec = Rate_per_Hr / Sec_per_Hr
(kps = 80 kph / 3,600 sec_per_hr = 1/45 = 0.022)

(952) Missing Second "A" Goal (Time): Time = K_Distance /- Rate_per_Sec
(Time = 45K sec)

(954) Missing First "B" Goal (Time): Time_Hr = K_Distance / Rate_per_Hr
(Time_hr = K / 80 kph)

(955) Missing Second "B" Goal (Time Convert): Time_Sec = Time_Hr * Sec_per_Hr
(Time_Sec = K/80 * 3,600 sec per hr = 45K)

(957) Missing First "C" Goal (Time): Sec-per-kilom = Hr-per-kilom * Sec-per-Hr
(Sec = 1/80 * 3,600 sph = 45 sec-per-kilom)

(958) Missing Second "C" Goal (Time Convert): Sec-per-K-kilom = Sec-per-kilom * K kilom
(Sec = 45 sec-per-kilom * K kilom = 45K)

Detailed Error Descriptions and Examples

This section provides more detailed descriptions of the errors listed on the "Bug Summary" pages.

Three canonical solutions, identical to those from the "Problem Statements and Correct Solutions" section, are repeated here for convenience. Examples of errors are shown as deviations from these canonical solutions. The solution from which the example deviates is indicated by DRT for distance problems, GR for graduated rate problems, and DRT-V for distance problems that use a variable in the solution. In most cases only the relevant modified lines are given, and they are numbered to match the canonical solution. For a number of examples, additional lines are provided for context. Lines on which the error occurs are marked by asterisks. In a few cases, intermediate steps that are not usually shown in the solution are presented in parentheses to clarify the example.

Canonical Solutions

(Roman Numerals Indicate Goals)

DRT (Five-Goal Problems)

DRT-05. A machine was set to produce the first 128 electrical components of the 420 needed for a certain production order at a rate of 24 components per hour and the rest of the order at a rate of 40 components per hour. If the machine started filling the order at 7:20 a.m. and ran continuously, at what time, to the nearest minute, was the order completed?

I. 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours

II. 3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components

III. 5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours

IV. 7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 12 hours 38 minutes

V. 9. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
10. Ending Time for Production = 7:58 p.m.

GR (Three-Goal Problems)

GR-13. A lawyer charges $100 for the first hour of service and $75 for each additional hour. A bill of $625 represents how many hours of the lawyer's service?

I. 1. Lawyer's Charge After First Hour = $625 - $100
2. Lawyer's Charge After First Hour = $525

II. 3. Number of Additional Hours = $525 / $75
4. Number of Additional Hours = 7 hours

III. 5. Total Number of Hours = 7 hours + 1 hour
6. Total Number of Hours = 8 hours

Canonical Solutions (cont.)

(Roman Numerals Indicate Goals)

DRT-V (Two-Goal Problems)

DRT-V-14. A car is traveling at an average speed of 80 kilometers per hour. On the average, how many seconds does it take the car to travel K kilometers?

Correct solution A

I. 1. Kilometers Traveled per Second = 80 kilometers per hour / 3,600 seconds per hour
   2. Kilometers Traveled per Second = 1/45 kilometers per second

II. 3. Seconds to Travel K Kilometers = K kilometers / 1/45 kilometers per second
   4. Seconds to Travel K Kilometers = 45K seconds

Correct solution B

I. 1. Hours to Travel K Kilometers = K kilometers / 80 kilometers per hour
   2. Hours to Travel K Kilometers = K/80 hours

II. 3. Seconds to Travel K Kilometers = K/80 hours * 3,600 seconds per hour
   4. Seconds to Travel K Kilometers = 45K seconds

Correct solution C

I. 1. Seconds per Kilometer = 1/80 hour per kilometer * 3,600 seconds per hour
   2. Seconds per Kilometer = 45 seconds per kilometer

II. 3. Seconds to Travel K Kilometers = 45 seconds per kilometer * K kilometers
   4. Seconds to Travel K Kilometers = 45K seconds

Detailed Error Descriptions

Math Bugs

Mathematical Errors

(109) Division Operator with Multiplication Operation, K / (1/45) = K/45
Although the operation is shown as division, a multiplication is executed.

DRT-V-A 3. Time in Seconds to Travel K Kilometers = K kilometers / 1/45 kilometers per second
4. Time in Seconds to Travel K Kilometers = K/45 seconds

(110) Tenths Match, Hundredths Do Not: 5.3 instead of 5.33

DRT 2. Production Time1 = 5.30 hours

(111) Units Match, Tenths Do Not: 7.2 instead of 7.3

Errors in the hundredths only (bug #110) are checked before errors in tenths only (this bug #111).

DRT 6. Production Time2 = 7.2 hours

(113) Decimal Shift: 53.3 instead of 5.33

The obtained value has a shifted decimal with respect to the expected value.

DRT 2. Production Time1 = 53.3 hours

(114) Remainder of Division Treated as a Decimal: 292/40 = 7 remainder 12 = 7.12 hours

(115) Computation Error, Not Identified by Other Math or Plan Errors: 292/40 = 6 [illegible]

This error is used for any equation in which a complex expression is set equal to an incorrect value that cannot be ascribed to a "Specific Plan Bug" or to a more specific "Math Bug."

(116) Decimal Treated as Fraction: 12.6 as 12 1/6

The decimal portion of a value is treated as a fractional value.

7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 12 1/63 hours

Specific Plan Bugs: DRT and DRT-V Time Bugs

(301) Division Remainder Treated as Time:

DRT 1. Production Time1 = 128 components / 24 components per hour
(Production Time1 = 5 remainder 8 hours)
*2. Production Time1 = 5 hours 8 minutes

(302) Decimal Portion Treated as Time:

DRT 1. Production Time1 = 128 components / 24 components per hour
(Production Time1 = 5.33 hours)
*2. Production Time1 = 5 hours 33 minutes

(304) Time Treated as a Decimal:

DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5 hours 20 minutes
*. Production Time1 = 5.20 hours

(305) Shift A.M. to P.M. or P.M. to A.M.:

This bug is sometimes triggered by student failure to indicate either a.m. or p.m., in which case a.m. is assumed.

DRT 11. Ending Time for Production = 7:58 a.m.

(306) Not Exact Match, but Within 1 Minute:

DRT 2. Production Time1 = 5.34 hours

(307) Use Decimal for Colon in Time:

A clock time is represented with a decimal in place of a colon.

DRT 10. Ending Time for Production = 7.58

(351) Convert to Minutes Instead of Seconds:

In making a time conversion the change is made to minutes instead of seconds.

DRT-V-A 1. Rate of Travel per Second = 80 kilometers per hour / 60 seconds per hour
Rate of Travel per Second = 1.33 kilometers per second

(352) Invert Time Conversion:

The rate is multiplied by the time conversion instead of divided by the conversion.

DRT-V-A 1. Rate of Travel per Second = 80 kilometers per hour * 3,600 seconds per hour
Rate of Travel per Second = 288,000 kilometers per second

(361) Time = Rate x Distance:

The incorrect formula T=RD is used in place of T=D/R.

DRT-V-A 2. Time in Seconds to Travel K Kilometers = K kilometers * 1/45 kilometer per second
Time in Seconds to Travel K Kilometers = K/45 seconds

(362) Time = Rate / Distance:

The incorrect formula T=R/D is used in place of T=D/R.

DRT-V-A 2. Time in Seconds to Travel K Kilometers =1/45 kilometers per second / K kilometers
Time in Seconds to Travel K Kilometers = 1/45K seconds

(363) Time = Rate / Distance:

The incorrect formula T=(1/R)/D is used in place of T=D/R.

DRT-V-A 2. Time in Seconds to Travel K Kilometers = (1 second / 0.022 kilometers) / K kilometers
Time in Seconds to Travel K Kilometers = 45/K seconds

Specific Plan Bugs: Other DRT and DRT-V Bugs

(401) Wrong Rate Value in a Structurally Correct Plan:

DRT 1. Production Time1 = 128 components / 44 components per hour

(402) Wrong Quantity Value in a Structurally Correct Plan:

DRT 3. Component Quantity2 = 450 components - 128 components

(404) Rate2 Used for Rate1:

DRT 1. Production Time1 = 128 components / 40 components per hour

(405) Rate1 Used for Rate2:

DRT 5. Production Time2 = 292 components / 24 components per hour

(406) Total Quantity Used for Partial Quantity:

DRT 5. Production Time2 = 420 components / 40 components per hour

(407) Use Partial Instead of Total Elapsed Time:

The ending time uses only one of the elapsed times instead of the total elapsed time.

DRT 10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes

(411) Unknown Value Used for Time1:

The structure of the solution suggests that Time1 is being used, but the source of the value used is unknown.

DRT *7. Total Production Time = 4 hours + 7.3 hours
8. Total Production Time = 4 hours + 7 hours 18 minutes
9. Total Production Time = 11 hours 18 minutes
10. Ending Time for Production = 7:20 a.m. + 11 hours 18 minutes
11. Ending Time for Production = 6:38 p.m.

(412) Unknown Value Used for Time2:

The structure of the solution suggests that Time2 is being used, but the source of the value used is unknown.

DRT *7. Total Production Time = 5.33 hours + 6 hours
8. Total Production Time = 5 hours 20 minutes + 6 hours
9. Total Production Time = 11 hours 20 minutes
10. Ending Time for Production = 7:20 a.m. + 11 hours 20 minutes
11. Ending Time for Production = 6:40 p.m.

(413) Unknown Value used for Total Time:

DRT 10. Ending Time for Production = 7:20 a.m. + 10 hours
11. Ending Time for Production = 7:58 p.m.

(414) Average Rate Not Weighted by Time:

An average rate is computed improperly since it is not weighted according to the different times. This example assumes a nonstandard approach to the problem.

DRT Average Rate = (24 components per hour + 40 components per hour)/2
Average Rate = 32 components per hour
Average Time = 420 components / 32 components per hour
Average Time = 13.12 hours

(415) Minute Total Treated as Increment:

DRT 9. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
(Ending Time for Production = 7:20 + 12 hours + 58 minutes)
10. Ending Time for Production = 8:18 p.m.

(416) Hours Treated as Minutes:

DRT 9. Ending Time for Production = 7:20 a.m. + 12.63 (hours)
10. Ending Time for Production = 7:33 a.m.

(417) Minutes Treated as Hours:

A time increment in minutes (e.g. DRT-07) is treated as an increment in hours.

DRT 6:15 p.m. + 5.38 (minutes) = 12:05 a.m.

(421) Treat Constant as 1:

The problem is solved as though the constant were equal to 1.

DRT-V-A 1. Rate of Travel per Second = 80 kilometers per hour / 3,600 seconds per hour
Rate of Travel per Second =1/45 kilometers per second

2. Time in Seconds to Travel K Kilometers = 1 kilometer/ 1 /45 kilometers per second
Time in Seconds to Travel K Kilometers = 45 seconds

General Plan Bugs

(701) Expression Not Reduced:

An expression is not sufficiently reduced. This bug is reported only if the nonreduced value is not resolved later in the solution.

DRT 8. Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
No subsequent lines in solution

(702) Final Goal Not Reduced:

The "final" answer to the problem is not reduced. It is like the no-reduction bug, but applies to the final goal.

DRT 10. Ending Time for Production = 7:20 a.m. + 12 hours 38 minutes
No subsequent lines in solution

(703) Multiplication Used Where Division Required:

DRT 1. Production Time1 = 128 components * 24 components per hour

(704) Division Used Where Multiplication Required:

%A 2. Commission per Set = 0.04 / $800 per set

(705) Addition Used Where Subtraction Required:

DRT 3. Component Quantity2 = 420 components + 128 components

(707) Correct Plan Structure, Unknown Values:

The student uses a value of unknown origin, but it fits into the problem solution structure for subsequent goals. Unlike bug #708, this bug is not restricted to a single value. The correctness of the plan is usually recognized because of the appropriate use in subsequent goals.

DRT *7. Total Production Time = 6.6 hours + 6.3 hours
8. Total Production Time = 6 hours 36 minutes + 6 hours 18 minutes
9. Total Production Time = 12 hours 54 minutes
10. Ending Time for Production = 7:20 a.m. + 12 hours 54 minutes
11. Ending Time for Production = 8:14 p.m.

(708) Correct Plan Structure, Single Unknown Value:

A plan has a single unexplained value within the correct structure. This is in contrast to bug #707 which includes errors in multiple values. This bug is used primarily in INTEREST and WORK problems. DRT problems use more specific bugs, 401 and 402.

WORK 3. Time for Desired Nondefective Bottles = 8.000 bottles / 65 bottles per minutes

(709) Label Right, Value Close; No Specific Explanation:

This error is used when the label is the only basis for identifying what the student is doing. It indicates that the combination of other plans and bugs could not adequately explain the value. At the same time, the value assigned to the bug must be within a reasonable range of the expected value, in this case within 1% of the larger of the obtained and expected values. Notice that this is not considered a math bug (cf #112) because there is no explicit statement of how the student derived the value.

DRT 3. absent
4. Component Quantity2 = 280 components

(710) Subtraction Used Where Addition Required: Total Production Time = 7.3 hours - 5.33 hours

As with other bugs of this general form, this bug should be identified only when it is seen as a transformation of an otherwise "correct" plan. If the values and operator are jointly wrong, then there is a missing goal.

DRT 9. Total Production Time = 7.3 hours - 5.33 hours

(711) Set Correct Response to Wrong Value: C/18 = 144

Although the correct expression for a goal is present in the solution (e.g. C/18 for problem DRT-V-17), that expression is set equal to an inappropriate value (e.g. 144).

DRT-V-A 4. 45K = 80

Specific Plan Bugs: Graduated Rate

(802) Use first rate as only rate

The rate for the first unit of time is assumed to be the rate for the entire time.

GR 1. Number of Hours = $625/ $100 per hour = 6.25 hours

(804) Rate Average as Hourly Cost

The difference between first unit and subsequent rates is taken as the average rate for the entire time period.

GR 1. Rate = $100 per hour - $75 per hour
Rate = $25 per hour
2. Time = $625 / $25 per hour
Time = 25 hours

(805) Rate Average as Hourly Cost

The average of the first unit rate and subsequent rate is taken as the average rate for the entire time period.

GR 1. Rate = ($100 per hour + $75 per hour) / 2
Rate = $87.50 per hour
2. Time = $625 / $87.50 per hour
Time = 7.14 hours

Missing Goals

These bugs indicate that the stated goal is missing. No goal can have any other associated bug if it is "missing."

DRT Missing Goals

(911) Missing First Goal, Time1 = Quantity1 / Rate1
DRT 3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
10. Ending Time for Production = 7:20 a.m. + 7 hours 18 minutes
11. Ending Time for Production = 2:38 p.m.

(912) Missing Second Goal, Quantity = Total Quantity - Quantity1
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
5. Production Time2 = 128 components / 40 components per hour
6. Production Time2 = 3.2 hours
7. Total Production Time = 5.33 hours + 3.2 hours
8. Total Production Time = 5 hours 20 minutes + 3 hours 12 minutes
9. Total Production Time = 8 hours 32 minutes
10. Ending Time for Production = 7:20 a.m. + 8 hours 32 minutes
11. Ending Time for Production = 3:52 p.m.

(913) Missing Third Goal, Time2 = Quantity2 / Rate2
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
11. Ending Time for Production = 12:40 p.m.

(914) Missing Fourth Goal, Total Time = Time1 + Time2
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
10. Ending Time for Production = 7:20 a.m. + 5 hours 20 minutes
11. Ending Time for Production = 12:40 p.m.

(915) Missing Fifth Goal, Finish Time = Start Time + Total Time
DRT 1. Production Time1 = 128 components / 24 components per hour
2. Production Time1 = 5.33 hours
3. Component Quantity2 = 420 components - 128 components
4. Component Quantity2 = 292 components
5. Production Time2 = 292 components / 40 components per hour
6. Production Time2 = 7.3 hours
7. Total Production Time = 5.33 hours + 7.3 hours
8. Total Production Time = 5 hours 20 minutes + 7 hours 18 minutes
9. Total Production Time = 12 hours 38 minutes

Graduated Rate M'ssing Goals

(941) Missing First Goal, Charge After First Unit of Time = Total Charge - Charge for First Unit of Time

2. Number of Additional Hours = \$625 / \$75
   Number of Additional Hours = 8.33 hours
3. Total Number of Hours = 8.33 hours +1 hour
   Total Number of Hours = 9.33 hours

(942) Missing Second Goal, Time at Subsequent Rate = Cost at Subsequent Rate / Subsequent Rate
This bug occurs only when #943 is also present.

1. Lawyer's Charge After First Hour = \$625 - \$100
   Lawyer's Charge After First Hour = \$525

(943) Missing Third Goal, Total Time = Time at Base Rate + Time at Subsequent Rate
Ignore the addition of the base unit (usually the charge for 1st hour, 1st mile, etc.)

1. Lawyer's Charge After First Hour = \$625-\$100
   Lawyer's Charge After First Hour = \$525
2. Number of Hours = \$525 / \$75
   Number of Hours = 7 hours

DRT-V Missing Goals

Form A:

(951) Missing First "A" Goal - Rate Convert,
Rate of Travel in Converted Units = Rate of Travel in Base Unit / Unit Conversion.

A 3. Seconds to Travel K Kilometers = K kilometers / 80 kilometers per second
4. Seconds to Travel K Kilometers = K/80 seconds

(952) Missing Second "A" Goal - Time,
Time to Travel Specified Distance = Distance to Travel / Rate of Travel.

A 1. Kilometers Traveled per Second = 80 kilometers per hour / 3,600 seconds per hour
2. Kilometers Traveled per Second - 1/45 kilometers per second

Form B:

(954) Missing First "B" Goal - Time,
Time to Travel in Base Units = Distance to Travel / Rate of Travel.
There is no time decomposition of Time in Base Units, so this error should normally occur only when the second goal is also missing.

(955) Missing Second "B" Goal - Time Convert,
Time to Travel in Converted Units = Time to Travel in Base Units / Unit Conversion.

B 1. Hours to Travel K Kilometers = K kilometers / 80 kilometers per hour
2. Hours to Travel K Kilometers = K/80 hours

Form C:

(957) Missing First "C" Goal - Time,
Time to Travel in Converted Units = Rated of Travel in Base Units * Conversion Units.

C 3. Seconds to Travel K Kilometers = 1/80 hour per kilometer * K kilometers
4. Seconds to Travel K Kilometers = K/80 hour

(958) Missing Second "C" Goal - Time Convert,
Time to Travel K kilometers in Converted Units = Rate of Travel in Converted Units * Distance

C 1. Seconds per Kilometer = 1/80 hour per kilometer * 3,600 seconds per hour
2. Seconds per Kilometer = 45 seconds per kilometer

54020-11044 • Y44P.7 • 209110