An Online Grammar-parsing Program That Decodes
Ordinary English Descriptions and Comparisons of Percentages and Rates

Tom Burnham[1] and Milo Schield[2]
Cognitive Consulting[1]
Augsburg College[2]
Dept. of Business Administration, 2211 Riverside Drive, Minneapolis, MN 55454

Abstract

To be statistically literate, students must be able to describe and compare rates and percentages in ordinary English based on data presented in tables and graphs. Since students require differing amounts of practice to acquire this skill, spending class time for drill takes time away from other important topics. An on-line grammar parsing program has been developed as part of the W. M. Keck Statistical Literacy Project. The current version handles 11 forms of descriptions and 11 forms of comparisons. At least half the students found the current program to be quite or very useful to them in learning the material. Limitations of the program are reviewed along with possible extensions.

1. Program Overview

The background for this program is reviewed.

1.1. Statistical Literacy

Statistical Literacy studies the use of statistics as evidence in everyday arguments (Schield 2004). These arguments are presented in ordinary English rather than in symbolic form. The statistics involved in these arguments are often ratios – rates and percentages – that are presented in tables and graphs. Students have considerable difficulty describing and comparing ratios (Schield 2000) when they are presented in tables and graphs (Schield 2001). It takes considerable class time to teach this material to students. Drilling on this is a burden on class time so a computer-aided drill program is needed.

1.2. Ordinary Language for Ratios

Statisticians have developed an algebraic notation for describing conditional probabilities that is clear, concise and unambiguous. In talking about adults, let M = Men and R = Runner, then P(R|M) is the prevalence of Runners among Men (the probability that a randomly selected man will be found to be a runner). Unfortunately, most students do not speak or even think in terms of algebra. Even if they do so for a course, they are unlikely to retain much if they don't use algebra thereafter. A great majority of students use ordinary English to describe and compare conditional probabilities in the forms of rates and percentages.

Schield (2000) organized ordinary language descriptions of ratios into four grammatical families: "% of,"

"percentage," "rate" and "chance." "Percentage" includes fraction and share. "Rate" includes prevalence and incidence. "Chance" includes risk, odds[1], likelihood and probability.[2] In making comparisons, there is also the "likely/prevalent" grammar.

A statistically-literate student should be able to decode any table of rates or percentages, describe any ratio using any applicable form of grammar and compare any two ratios using any applicable grammar.

Table 1: Unemployment Rates in 1997 (% of civilian labor force) by Educational Attainment, Sex and Race

| SEX and RACE | Total | Highest Grade Achieved | | | |
		< 12	12	13-15	> 15
Total: \2	4.4	10.4	5.1	3.8	2.0
Male:	4.7	9.9	5.6	4.0	2.1
Female:	4.1	11.3	4.5	3.6	2.0
White:	3.9	9.4	4.6	3.4	1.8
Black:	8.1	16.6	8.2	6.1	4.4
Hispanic: \4	7.3	9.6	7.5	5.5	3.0

\2 Includes other races, not shown separately.
\4 Persons of Hispanic origin may be of any race.

The margin values in the top row are averages so the rows are wholes. The margin values in the left column are averages so the columns must be wholes. The data are all ratios so there must be a part and so it must be in the title. In this case, "Unemployment" is the part common to all ratios while the "civilian labor force" following "% of" is a whole.

In referring to the 1997 US civilian labor force in Table 1, the circled 4.7% can be described 11 ways:

The "% of" grammar involves two forms:
1. 4.7% of males are <u>unemployed</u>
2. Among men, 4.7% are <u>unemployed</u>.

"Percentage" grammar involves two forms:
1. The percentage of men who are <u>unemployed</u> is 4.7%
2. Among men, the percentage of <u>unemployed</u> is 4.7%.

"Rate" grammar involves seven forms:
1. The rate of <u>unemployment</u> is 47 per 1,000 men.
2. The <u>unemployment</u> rate of men is 47 per 1,000.
3. Men's <u>unemployment</u> rate is 47 per 1,000.
4. Men are <u>unemployed</u> at a rate of 4.7%.
5. <u>Unemployment</u> occurs at a rate of 47 per 1,000 men.

---

[1] The grammar for odds is unique within this family.
[2] The "chance" family also includes confident and accurate. E.g., 95% confident, 90% accurate.

6. The rate at which men are <u>unemployed</u> is 4.7%.
7. The rate at which <u>unemployment</u> occurs is 47 per 1,000 men.

This program handles these 11 forms of descriptions but is not programmed to handle "chance" grammar. Each description involving "percentage" or "rate" has a corresponding comparison. While there are no comparisons using "% of" grammar, there are two forms involving "likely/prevalent" grammar for a total of 11 kinds of comparisons. Most of these have both a long and a short form, and all of these can involve one of three comparisons.[3]

1.3. Need for an Online Program

For those who did not learn English at an early age, describing and comparing ratios in ordinary English involves learning the rules for using prepositions and relative clauses. Yet those who have learned English as a second language may do better than native English speakers since the native speakers are not accustomed to using English in such a precise manner. Both groups need time to practice decoding tables and writing descriptions and comparisons of conditional probabilities in ordinary English.

In teaching this material to over a 1,000 students over a period of 10 years, it seemed that students needed 3 to 6 hours of class time for drill for 80% of the class to achieve a passing level of skill at describing and comparing both percentages and rates in tables. This is a lot of class time. Yet 20% of the students needed still more time. Tutors were provided but busy students have difficulty working in special study sessions. For these reasons an on-line grammar-parsing drill program could be valuable to both students and teachers.

It may be that asking a computer program to decode ordinary language statements is asking too much. Great minds working in artificial intelligence have realized that analyzing context is not easy for computers. But this situation differs in several ways. First, the semantics of all correct answers is known in advance. Second, the patterns that can be used to describe or compare rates and percentages are fairly limited in their grammatical structures. Third, the semantic domain for answers is limited. These facts make it possible to create a program that is useful to students.

If an on-line program simply says OK or Wrong, this is not helpful to a student. A teacher or tutor would describe the error in such a way as to help the student without giving them a correct answer. The quality of the program will be largely determined by the appropriateness and helpfulness of the program error messages.

1.4. Goals

The program goal is to help users achieve the user goal.

- User Goal: To write a single sentence in ordinary English that describes a single ratio or compares two ratios contained in a table of percentages or rates and that can be quoted out of context without being misunderstood (no grammar errors, no ambiguities and no missing relevant terms).
- Program Goal: To decode the user's sentence describing a ratio or comparing two ratios contained in a table of percentages or rates, to accurately identify user errors relative to the table data and the problem statement, to provide error messages that help the user improve their answer, and to capture student comments for program improvement.

1.5. Program Overview

This program was designed in Visual Basic for two reasons. Visual Basic (VB) handles screen input and output with little overhead and VB may be readily converted into on-line active server pages (ASP) for web deployment. The first author[4] designed and programmed the system. The second author supplied the general specifications and the test data, and tested the system with students. Both authors had worked together in developing descriptive rules for describing and comparing ratios in various grammatical styles.

The basic tension in designing this system is to avoid making it so specific that each problem stands alone as a separate program, and yet avoid making the program so general that it must be able to decode all ordinary language statements. Finding a balance between specificity and generality is an extremely difficult matter.

1.6. Program Specification

These are the general specifications for the program to validate statements describing and comparing ratios.
- Present problem statement and data table.
- Accept user input
- Analyze misspelled or extraneous words.
- Analyze semantic errors in part-whole, test-base, pattern or amount.
- Provide appropriate feedback to the user.
- Store a log of each session.

The program design is not that of artificial intelligence (AI) nor is it just a right-answer lookup. The program handles some 22 core patterns for descriptions (11) and

---

[3] 1. Simple Difference: {Test} is X more than {base}. E.g., 8 is 6 more than 2. 2. Simple Ratio: {Test} is X times as much as {base}. E.g., 8 is 4 times as much as 2. 3. Relative Difference. Percentage Compare: {Test} is X% more than {base}. E.g., 8 is 300% more than 2. Times Difference: {Test} is X times more than {base}. E.g., 8 is 3 times more than 2. The program does not handle #1.

[4] Tom Burnham, the copyright holder of this program, was a senior systems designer at DataPoint where he worked on office automation software, run-time packages, utilities and operating systems.

comparisons (11) each having many variations. The program design calls for a computer-aided drill – not for a computer aided tutorial. Students must know the rules for the various grammars.

A key feature of the system design is the user log. Every time a user accesses the system, a unique user log is opened. This log contains the question being asked, the user response, the program-generated answers and any user comments. User logs generate one-line entries in a summary log. These logs are reviewed to see what program changes might be helpful.

The program uses a general dictionary plus a problem dictionary for each problem. The general dictionary contains over a 100 pairs of entries. Each pair maps the first word (a generic word in the user input) into the second word (a general semantic category where the category name may be one of the members).[5] The problem dictionary contains words specific to the problem (those in the table and the question and reasonable synonyms) and indicates for each whether it is part or whole, test or base

2. Analyzing Input

Analyzing the input is a major task of any program that attempts to read ordinary English. It must deal with extraneous or irrelevant terms, synonyms, equivalent forms, repeated words, omitted words, modifiers following the part and intervals (e.g., age ranges).

2.1. Extraneous or Irrelevant Terms

User input may involve extraneous terms: terms that do not appear in the table. Extraneous terms include "those surveyed," "respondents," "people," "distributed by," "according to" and "those who answered." Using extraneous terms may reflect a need to indicate a common whole when it is not stated. This creates problems for the program since there are many possibilities. Each one must be included in the problem dictionary or the user gets the message "XXX not in dictionary."

User input may involve irrelevant terms: words appearing in the data table (e.g., "by Sex") that are not relevant to describing a ratio (e.g., use "male" or "female").

2.2. Synonyms

The use of synonyms is common in English. But synonyms create major problems for an automated program. Synonyms can involve a word-for-word equivalent, or they can involve a semantic equivalent involving a different phrase or clause.

Word-for-word equivalents are common. "Male" and "man" may be equivalent when talking about adults. Other equivalents in a given context include "baby" and "infant," "teenager" and "adolescent," "senior" and "elderly," "worker" and "employee," "manager" and "boss ," and "smart" and "intelligent."

English is an obstreperous language. There are many ways to use syntax to indicate a given semantic idea. Consider families who have cats. One can say "families who own cats," "cat-owning families," "cat own-ers," "the owner of a cat," etc. Yet specific grammar is needed to indicate the part-whole status of words. Using different phrases to express the same meaning makes parsing much more difficulty. The shift from a single-word adjective form to a trailing prepositional phrase is not uncommon. For example in place of "city schools" a student might write "schools in the city."

Equivalent forms can involve a change in the predicate. A common case is the change from a copula (is/are) to a verb (transitive or intransitive). Consider the shift from "20% of kids are drug users" (copula) to "20% of kids use drugs" or "20% of kids do drugs."

2.3. Equivalent Forms

English allows many ways to say the same thing. The flexibility of English makes it much more difficult to develop a program to handle such matters. In ordinary English, one can sometimes reverse the subject and predicate. Thus, X% of {Whole} are {part} could be rewritten as, {Part} are X% of {Whole}. Since a few students used these forms, the current program handles these inversions for "percentage" grammar and for the first three forms of "rate" grammar.

The use of possessives in "rate" grammar is common. E.g., "The unemployment rate of blacks" can be rewritten as "The black's unemployment rate." The current program handles the possessive form. Dropping the possessive indicator (the apostrophe) is unfortunately all too common in ordinary use. E.g., "the black unemployment rate." Decoding such usages requires content knowledge and so these usages are not allowed.

One synonym problem involves the use of paired words or phrases. To handle this, the program should be a phrase-based system. E.g., "less than 12 years' education" is the same as "no high school diploma."

2.4. Repeated Words

Sometimes the same word appears in both part and whole in a description or in part, common whole, test whole and base whole in a common-part compare. The current program cannot properly handle multiple instances of the same term.

---

[5] At this point, there are 26 semantic categories each containing a single word (e.g., "among" and "than" are separate one-word categories) and there are 24 semantic categories each involving multiple words (e.g., the "percent" category includes both "percent" and "%").

Consider using "likely" grammar to compare public elementary schools and public high schools on the percentage that reported a serious violent incident. One might legitimately say, "Among public schools, the percentage that reported a violent incidence is higher among high schools than among elementary schools." Note the repeated use of "schools." In this example, it is present in the common whole (among public schools), in the test whole (among high schools) and in the base whole (among elementary schools).

2.5. Word Omission in Comparisons

Word omission is a common problem. It may be justifiable when the context (the content of the words) provides information on the missing word or phrase. For example, instead of saying "Six is 3 times as much as two" we say, "Six is 3 times two."

Word omission may make it impossible to determine whether a comparison involves a common part or two distinct parts. Consider this comparison, "Accidental death is more likely among men than women." From content we know that "men " and "women " should be treated as linked – as having the same part-whole status. Since "men" is a whole, then so is "women" and the comparison is a common-part comparison with "accidental death " as the common part in two ratios. But a grammar parser doesn't know this. The grammar parser sees, "P is more likely among W than Z." Z could be linked to the part in a distinct-part comparison: "P is more likely among W than Z *is among W*." While common-part comparisons are much more common than distinct part comparisons, a grammar parser cannot be expected to rely on this so the current program requires a user to enter "among" twice in this example.

2.6. 'Among' Following a Part

"Among" typically indicates a whole, but it sometimes follows a part – which may create confusion. For example, suppose that 10% of men are unemployed among blacks. If blacks modified unemployed, then blacks would be part. The current program treats this trailing phrase in a description as modifying the entire statement so that among introduces a whole.

2.7. Modifiers Following a Part

The preceding case where "among" follows a part exemplifies a more general problem: the use of other words or phrases following a part which may or may not indicate a whole depending on the content of the substantive term. One particular problem involves modifiers following "the rate of {part}." Consider this:

The rate "**of death** _____ **X**" is # per 1,000 W.
- of death among X   (X is whole/denominator)
- of death due to X   (X is part/numerator)
- of death for/in X   (X is ambiguous)

The program requires that non-standard prepositions (prepositions other than "among," "of" and "per") indicating wholes be placed at the start of a description to indicate that they apply to the entire statement rather than being placed after the part where their part-whole status is grammatically ambiguous.

2.8. Decoding Intervals

The use of intervals to describe groups of data is common for ages, incomes and other numerical measures. But there are a many ways to describe an interval in English. In the case of one interval, 15 students created 18 ways of describing the interval – none of which were alike. This creates a major problem for a program to recognize all the phrases students may use in describing these intervals.

Closed intervals seem to have fewer variations than open intervals. And top open intervals seem to have fewer variations than bottom open intervals. And bottom intervals are more likely to include comparison words (less than) that are top open intervals. Top intervals are more likely to be described using "65 and above" or "65 and up" rather than "65 and greater." The current program decodes only age intervals.

2.9. Field Testing and Test Data

As indicated in the previous sections, there are many ways to express the same semantic relation using ordinary English syntax. There is no way to anticipate all possible variations. A new problem must first be field tested with students to see what synonyms are most common. In our experience, the results from at least 20 students are needed to serve as an adequate base for developing a dictionary for describing a given ratio in a table or for comparing two ratios in a table. Monitoring student answers and comments in the on-line error log is absolutely essential to identify synonyms that should be added to the dictionary. Otherwise students become unhappy and loose motivation when they give a correct answer but the program says, "Word not in dictionary."

2.10. Program Design

This program is not a pattern matching game. It does real structural analysis using a parser like that found in many compilers. This allows the program to handle a much wider range of syntactic combinations. But it makes the program logic considerably more complex.

The current program tries to follow the logic of the sentence first and compare the resulting semantics with the correct answer. But if the respondent omits key information, the program may be unable to determine what the respondent intended. Students would sometimes leave out "times" in comparisons or "more" in comparisons, leave out "percentage," "rate" or "likely" in comparisons, write horrible combinations

(such as "the percent of"), and omit the main verb (e.g., "Males 3 times as likely female.") It is very difficult for the program to identify the error in the respondent's thinking when key terms are omitted.

The opposite problem is when students combine elements from different grammars such as using both "percentage" and "likely" in a single comparison. E.g., "The percentage of men is more likely in the military."

When checking the value of an arithmetic comparison, the rule is that the numeric value must be within 5% of the target amount (to reduce rounding errors).

2.11. Program Limitations

As listeners, we have limits on how many different ways we interpret information. So does any program. Words that have two different grammatical roles are fertile sources for limitations. "Reported" can be either a verb or a participle (adjective) as in "the percentage that reported a violent incident" (verb) or "the percentage of reported violent incidents" (participle).

The current program does not use a general spelling dictionary.[6] The program does not analyze distinct part comparisons. E.g., "In the military, men are more likely than are women." The program can't readily handle words that are repeated in different contexts, so analyzing "less than" for both an interval and for a comparison is a difficult matter. The program does not handle forms beginning with "It is" such as "It is X% more likely that men will die accidentally than will women."

3. Giving Feedback

This program is not a computer-assisted instructor or a computerized tutor. It is designed to be a computer-assisted drill program. Users must be generally familiar with the syntax appropriate for each of the different grammars. Nevertheless, the utility of the program is largely determined by the quality of the feedback.

3.1. Terms

The following words and phrases are used by the program to analyze user answers:
- *Copula*: A connecting verb such as "is," "are," "was" or "were."
- *Whole & Part*: These are the pieces of a ratio presented as a pie.[7]
- *Slot*: Place in the Ratio Describe template where the part or whole is inserted.[7]

"*Word in wrong slot*" may mean "part" word in "whole" slot or vice versa. If a word is not misspelled, "*Word not in dictionary*" means that word is not in the problem dictionary so the user should try another word.

In writing comparisons, certain words and phrases have been used as follows to analyze user answers:
- *Test & Base*: "Base" is the basis of the comparison; "test" is the item being compared to the "base."
- *Slot*: Place in a ratio comparison template where the "test" or "base" is inserted.[7]

"*Word in wrong slot*" in a comparison can mean "base" word in "test" slot or vice versa.

3.2. Classes of Errors

Two major classes of errors are "slot errors" and "pattern errors." "Slot errors" involve a word in the wrong slot. Slots can involve either part-whole terms or test-base terms. Pattern errors are more abstract. The program does comment on style. "The percentage of TW who are P is greater than the percentage of P among BW" is correct but generates a "lack of parallelism" message.

4. Results

This program is currently at www.StatLit.org[8]. The use of this program is open to all and is encouraged. Teachers and data professionals are asked to use www.StatLit.org/RSVP. Other entry points are available for students in classes. All users should study the operating instructions and the templates for describing and comparing ratios before using the program.

4.1. Adding Problems to the Tutor

Faculty need to be able to add new tables and graphs to the tutor. At one level, this is a trivial matter. A new problem is created by adding a picture of a table or graph, a new question and populating the associated dictionary with the relevant terms in the problem. But this basic setup means that users can not use any synonyms. The next level is to identify an adequate range of word-for-word equivalent synonyms but this is difficult without testing.

4.2. Student Preparation

The success of the program is largely determined by the preparation students receive before using the program.[9] Students need training on
1. the grammar used in describing and comparing

---

[6] Users misspell keywords: "like" for "likely", "mor" for "more" and "time" for "times." They abbreviate: "pctg" for "percentage," "likelier" for "more likely", "bus" for "business," etc. These misspellings and abbreviations cause problems for the program and for the user.

[7] See templates for describing and comparing ratios in Schield (2000b) or in program splash screen.

[8] The VB program first ran on the web on 16 Feb, 2005.

[9] Weaker students may treat the program like a game. In a game if the clues are not obvious, one simply reverts to trial and error. This may be productive in a game, but it is not productive of user learning. Students who expect the program to be a computer-aided instructor or tutorial may find the learning time to be longer and the program to be less helpful than expected

2. the terms used by the program
3. how to get started with the program.
4. how to interpret program responses
5. how to get additional help from program
6. what to do given a particular program response.

All of this is necessary to minimize disappointment and frustration on the part of students. But items 2-6 take less class time than the needed drill!

4.3. Student Feedback

The ultimate test of any automated drill program is the student perception of the value of the responses. This on-line grammar checker has been used by some 70 students as of 30 June 2005 The 14 adult business majors who used the most recent version of the program were asked two questions using the following multiple choice answers: 0. Very unhelpful, 1. somewhat unhelpful, 2. OK, 3. Somewhat helpful and 4. Very Helpful. B indicates no answer.

"*How helpful is the program to you?*" 2.7 average: 44, 333333, 2222, 11, 0. Roughly, 50% find it helpful, 25% do not and 25% find it OK.

"*If the screen design, the questions and the answers were at their best, how helpful could this program be?*" 2.7 average: 4444, 33333333, 1, 00. Roughly, 75% believe the program could be helpful while 25% do not.

Note the three students at the low end in each of these questions. Although the survey was administered without any names, students gave in class feedback which is the basis for the following conclusion. There were three students who spent minimal time on the grammar templates. These students indicated frustration when the program failed to tutor them in improving their skills. The log files indicated these students used the program much less than the other students. This has implications for teaching and for program development.

4.4. Program Outcomes

The goal of this project was to design and program a computer-aided drill program that would be useful to students in learning how to use ordinary English to describe and compare rates and percentages presented in tables. Since 50% of the students found the current program to be helpful, we believe that this program has achieved the program design goal.

Using this program, future teachers of statistical literacy need not spend as much class time on such drill and statistical literacy can be offered on-line.

This program has been used by students in an on-line statistical literacy course (MAT 2050) offered by Capella University. Student comments include "a cool tool", "an invaluable tool", "a great help," "a great resource and very practical." and "extremely helpful."

4.5. Future Activities

This test system was designed to see if an on-line computer-aided drill program was feasible. Future enhancements include:

1. Analyze the input as phrases or clauses rather than just analyzing word patterns. This would dramatically expand the range of real-world problems that can be included. It would dramatically improve the usefulness of certain classes of error messages.
2. Handle other intervals (e.g., income or education).
3. Handle prevalence or incidence in "rate" grammar.
4. Handle fraction or share in "percentage" grammar.
5. Add "chance/risk/probability/likelihood" grammar
6. Add distinct part comparisons ("In the military, men are much more likely than are women").
7. Add program control of what table cell to analyze.
8. Add ability to display graphs
9. Add analyze comparisons of pure numbers.
10. Handle specification of the type of comparison.
11. Automate the generation of a problem dictionary given two descriptions of ratios to be compared.
12. Send results to pre-specified e-mail address.

As noted previously, this program is not just a ratio statement validator; it is a grammar parser. As such, it can be extended to a much wider range of problems where the rules for handling grammar are fairly well established. It might be extended to handle student statements about conditional relationships such as those describing confidence intervals and hypothesis tests.

Acknowledgments

References

Schield, Milo (1999). *Common Errors in Forming Arithmetic Comparisons. Of Significance* journal. The Association of Public Data Users, Volume 1. "1999SchieldAPDU.pdf"

Schield, Milo (2000a). *Statistical Literacy and Mathematical Reasoning*. International Conference on Mathematics Education. "2000SchieldICME.pdf"

Schield, Milo (2000b). *Statistical Literacy: Difficulties in Describing and Comparing Rates and Percentages.* ASA 2000 Proceedings of the Section on Statistical Education, p. 76. "2000SchieldASA.pdf"

Schield, Milo (2001). *Statistical Literacy: Reading Tables of Rates and Percentages.* ASA Proceedings of Statistical Education Section. "2001SchieldASA.pdf"

Schield, Milo (2004). *Statistical Literacy Curriculum Design.* 2004 IASE Round table on Curriculum Design, Lund Sweden. "2004SchieldIASE.pdf"
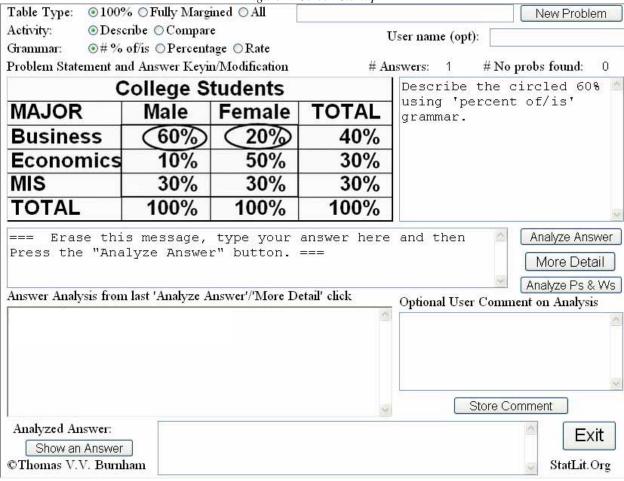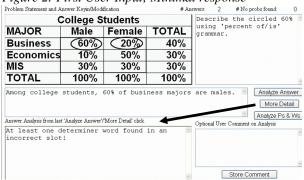
Appendix A. Screen Layouts

Figure 1: Screen Startup

Figure 2: First User Input; Minimal response
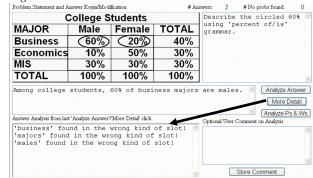

Figure 3: More Detail Results #1

Appendix B: Dictionary Setup

While there is an overall system dictionary with key words like "percentage" and "rate," each problem has a dictionary for the words that are unique to a given table. The quality of the program response is determined in large part by the quality of that problem's dictionary.

To set up a dictionary, enter the words that identify the referents of the ratio(s) in the table whether they are explicitly mentioned in the table or are word-for-word synonyms. Identify each word as a part, test whole, base whole or common whole. Since the current system only handles common-part comparisons, there is no need to distinguish common part from distinct part.

The current system does allow a word to be repeated provided it occurs in a different part of speech but not if it is repeated in two different part-whole categories.

2209

Appendix C: Error Messages

Dictionary errors must be fixed first. 'XXX' is a word from the user input. The brackets [..] enclose optional text; the braces {..} enclose alternate text.

DICTIONARY ERRORS:

1 'XXX' not found in dictionary!
2 No more checking until all dictionary errors are fixed!

GENERAL GRAMMAR ERRORS:

3 No ratio keyword found!
4 Too many ratio keywords!
5 The {word/phrase} 'XXX' must be preceded or followed by a copula.
6 The words 'XXX' must contain a comma to separate the (leading) common whole terms from the rest.
7 'XXX' is disallowed relative pronoun; only one is allowed in each percentage structure!
8 [#th occurrence of] 'XXX' is a disallowed redundancy!
9 'Percent' should be 'percentage'.
10 'rates' should be 'rate'.
11 'XXX' or preceding comma doesn't fit {some grammar}!
12 'XXX' doesn't fit {some grammar}!
13 Incomplete; end of answer before end of {some grammar} pattern!
14 No more checking until all pattern errors are fixed!
15 Answer is not in requested form ({some} grammar)!

SPECIFIC GRAMMAR ERRORS:

16 At least one determiner word is missing!
17 {word/phrase} 'XXX' missing!
18 {part/whole} {word/phrase} 'XXX' missing!

19 At least one determiner word found in an incorrect slot!
20 {word/phrase} 'XXX' found in wrong kind of slot!
21 {part/whole}{word/phrase} 'XXX' found{whole/part}slot

22 'XXX' cannot be negated for this problem!
23 'XXX' not relevant to problem!
24 # is incorrect value!
25 The phrase 'XXX' does not describe a margin cell of the current table!

"RATE" GRAMMAR ONLY:

26 # per # is incorrect rate!
27 Number of parts is missing!
28 Number of wholes is missing!
29 Number of wholes is zero, which is invalid!

COMPARISONS ONLY

30 'as' should be 'than'.
31 'than' should be 'as'.
32 Compare unit '... ' is incorrect!
33 'points' is not allowed in ratio comparisons!
34 [#th occurrence of] 'XXX' and [#th occurrence of] 'YYY' are not in parallel structures!
35 'XXX' missing between [#th occurrence of] 'YYY' and [#th occurrence of] 'ZZZ'!

FINAL MESSAGE:

36 No problems found.

Appendix D: Student Answers to Survey

Ten different versions of the program were tested with four groups of students (15 to 25 in each). Their feedback resulted in program improvements. The following feedback was obtained from students who dealt with the latest version. They were surveyed 25 June, 2005.

MODAL ANSWERS

Q1. Questions worked to date? 31 - 99

Q2. Answers entered before feeling comfortable? 11-20

Q3. Tables done before feeling comfortable? 4 - 5

Ordinal answers were from 5 choices with two positive, two negative and one neutral (OK).

Q4. How helpful is current program? Quite helpful.

Q5. How helpful would it be at its best? Quite helpful.

Q6. Did vocabulary conflict with class? Sometimes

Q7. Was program easy to learn? OK and Very easy.

Q8. Is number of tables OK? Not at all, not very & OK.

Q9. Is range (variety) of tables adequate? OK.

Q10. What needs most improvement? Dictionary.

Q11. Concerning the screen design, what needs most improvement? Indication of what to do when.

Q12. How clear are questions? Adequate, Clear.

Q13. How helpful were error messages? OK.

Q14. What answer needs most improvement? "Slot"

Q15. How helpful was "More Detail" button? OK.

Q16. How helpful was "Show Answer" button? Very.

Q17. How strongly would you recommend program to others? Moderately, Strongly and Very Strongly.

Appendix E: Student Comments

#2: LIKE Great for practice. Helped me understand how to read tables. DISLIKE/IMPROVE: Better instructions up front on how to use. Maybe more tables as examples.

#4. LIKE: I like [it] that the program … guides us through the phrases to … get to a correct answer.

#6. LIKE: Once I got the program [figured out], it was very helpful in getting all the details right on my answers (making sure I didn't leave out any words). IMPROVE: I just think it needs more variety -- more questions and tables.

#8. LIKE Innovative way to do homework and study. DISLIKE: Not enough tables.

#11 LIKE: How you could type the answer in and then see the correct one if wrong. DISLIKE: Hard to understand where to type at first using it, but got used to it eventually.

#14: LIKE: Good practice for quizzes. DISLIKE: "Not found in dictionary", [disallows] redundancy of words.