

WHAT ITEM WRITERS THINK WHEN WRITING ITEMS: TOWARDS A THEORY OF ITEM WRITING EXPERTISE

Dennis Fulkerson

Paul Nichols

David Mittelholtz

Pearson

For presentation at the American Educational Research Association Annual Meeting

Denver, Colorado

May 3, 2010



*Using assessment
and research to
promote learning*

With the increase in interest and use of innovative item types on large-scale assessments, test developers are experiencing pressure to quickly, efficiently, and cost effectively produce quality innovative items. The efficient development of quality innovative items is sometimes hindered by inexperienced item writers who are not familiar with the challenges and nuances of innovative item types. The study of expert item writers offers the possibility of “bottling” the knowledge and skills acquired by these experts over years of hard work. The descriptions of the identified conceptual knowledge and skills of expert item writers could be incorporated into item writing workshops in order to equip new item writers with the tools necessary to produce quality figural response items.

One type of innovative item is computer-delivered, scenario-based tasks. Scenarios consist of several related scenes that emphasize inquiry-based learning theory and hands-on science strategies and provide students opportunities to observe a process of science and the results of an investigation or event. Benchmark-aligned items are presented within the scenario context. Storyboards are products of innovative test development processes that precede the development of online scenarios. A storyboard is a written description of the narrative, images, animation, and/or video that will be developed for a test scenario. Figure 1 shows excerpts from an example five-scene science storyboard (Fulkerson, Nichols, Haynie, & Mislevy, 2009).

Storyboard Title: Snapping Turtles Life Science
Scene 1
Script Text: Snapping turtles are commonly found in ponds and rivers. Snapping turtles are characterized by webbed feet, hard shells, and sharp mouths.
Art Description: Art consists of an unlabeled still picture of a common snapping turtle. Listed characteristics are clearly displayed.
Scene 2
Script Text: Snapping turtles eat fish, invertebrates, amphibians, and carrion. Adult snapping turtles have few natural predators.
Art Description: Art shows a still picture of an adult snapping turtle with a frog in its mouth.
Scene 3
Script Text: Snapping turtles reproduce sexually. Female snapping turtles lay dozens of eggs at a time. The eggs are laid in sandy soil away from the water's edge. Offspring hatch underground and dig to the surface.
Art Description: Art shows a series of stills: 1) Adult snapping turtles in copulation, 2) Female snapping turtle laying eggs in a hole in sandy soil, 3) Young snapping turtles emerging from sandy soil.
Scene 4
Script Text: There are two species of snapping turtles in the United States. One species, the common snapping turtle, has many subspecies.
Art Description: A geographic map of North America showing the labeled ranges of the two snapping turtle species.

Figure 1. Excerpts from an example science storyboard (Fulkerson, Nichols, Haynie, & Mislevy, 2009).

A cognitive model of item writing may be particularly helpful in improving the quality of innovative items. However, cognitive modeling with respect to items has been limited to studying test takers' response processes and knowledge structures. This research has overlooked the item writers themselves. No effort has been made to construct cognitive models of item writers' writing processes and knowledge structures. Careful examination of item writers' processes and structures may facilitate efforts to improve the quality of items at an early phase of development by addressing and resolving areas of need in item writers' knowledge and skills related to item construction.

In an earlier study of experienced item writers' cognition, Fulkerson, Mittelholtz, and Nichols (2009) found that expert item writers engaged in three phases of problem solving. In the initial representation phase, expert item writers routinely created a mental model (e.g., Gentner & Stevens, 1983; Johnson-Laird, 1983) of the situation described in the scene. Problem definition statements were used to capture experts' creation of mental models. Problem definition predominantly occurred following the receipt and initial reading of the item writing assignment.

In the second phase, the exploration phase, the item writer purposefully explores the problem solving space in search of content that represents a workable solution to the assignment. Operator statements were used to capture experts moving toward a solution to the assignment. This phase involved a sustained sequence of processes.

In the third phase, the solution phase, the item writer successfully completes the assignment by finding a workable solution that satisfies the set of constraints. Solution statements were used to capture item writers achieving some desired situation such as completing the item writing assignment or achieving a subgoal eventually leading to completing the assignment. The solution occurred following a series of operators, toward the end of a sustained problem solving effort.

This study extends the earlier work of Fulkerson, Mittelholtz, and Nichols (2009) in two ways. First, this study includes writers with varying degrees of item writing experience. The inclusion of both inexperienced and experienced item writers allows the identification of thought processes, knowledge, and skills that may be acquired over years. Second, this study asked item writers to develop a storyboard that would support the development of a scenario as well as write an item for one storyboard scene. The

findings from this study extends the earlier item writing model of Fulkerson, Mittelholtz, and Nichols (2009) to include the development of innovative, scenario-based tasks.

Theoretical Perspective

This study has adopted the theory of insight problem solving as an initial guide to analyzing the study data. Insight problem solving is characterized by an initial period of purposeful problem solving activity that may result in the rapid completion of the solution. But sometimes the item writer experiences an impasse, a state of mind in which the item writer feels that all options have been explored and he or she cannot think of what to do next. The item writer's continued concentration on the problem often causes a new idea or option to come to mind. This so-called aha-experience is typically unanticipated by the item writer and is followed by rapid progress until the next impasse is encountered or the item is drafted.

Protocol analysis techniques (Ericsson & Simon, 1993) can be applied to explore how experts perform tasks such as writing items. Verbal reports of item writers' thinking while writing test items contain information on the knowledge, strategies, and reasoning employed by the writers as they develop items. Verbal reports offer an important tool in examining how item writers develop figural response items because they provide different, more direct evidence of item writers' thinking than do other methods such as observation and post-event surveys.

Study Methods

The participants in this study were three science content specialists from an assessment company. The participants had been employed as science content specialists for at least 4 years and had a range of prior teaching experience from 4 to 31 years. One writer had no experience in writing scenarios, while the remaining two writers had some experience writing scenarios over a three-year period. All participants had at least a master's degree. Demographic information for all four participants is shown in Table 1.

Table 1. Participant Demographic Information

Writer	Assessment Company Experience	Scenario Writing Experience	Teaching Experience	Subject Area	Highest Degree
1	4 years	3 years	24 years	Chem/Phys	Master
2	5 years	3 years	31 years	Chemistry	Master
3	6 years	None	4 years	Biology	Master

A training session was held four days before the first think aloud session was conducted. The training session required approximately one hour. All participants received information on the purpose of the study, the science content framework, and the task they would be asked to do.

The participants were tested individually in one-hour sessions. During individual sessions, participants received a writing assignment (Appendix A) and instructions for completing the task (Appendix B). The writing assignment presented three sets of assigned content benchmarks and asked participants to select one of the sets of benchmarks as their assignment. Benchmarks are the content objectives to which test tasks are aligned, and were selected from the *Minnesota Comprehensive Assessments Series II (MCA-II) Test Specifications for Science* (MDE, 2008).

The writing assignment asked participants to complete two tasks. First, participants were asked to write a rough 4-scene storyboard based on the assigned benchmarks. Second, participants were asked to write one rough multiple-choice item aligned to one of the assigned benchmarks in the context of one of the previously storyboard scenes. In addition to the writing assignment, participants received a storyboard scene template (Appendix C), a multiple-choice item template (Appendix D), and a copy of the Minnesota MCA-II Science benchmarks (MDE, 2008). As they responded to the writing assignment, writers were asked to think aloud, verbalizing cognitive information generated during item writing. Subject matter experts' verbal reports were audio recorded as they attempted to write items for each selected benchmark.

Data Analysis

The analyses occurred in four steps. First, the verbal behaviors recorded during the think alouds and retrospective reports were transcribed for analysis. Second, the transcripts were reviewed and edited for accuracy. Third, the transcripts were segmented into individual statements. Finally, each statement segment was coded into broad, general categories of problem solving. The coding categories are shown in Table 2.

Table 2. Labels and Descriptions of Statement Coding Categories

Statement Category	Description
Extraneous	Represent statements which appear irrelevant to the assignment
Nonconforming	Represent statements relevant to the assignment that do not fit any current coding categories.
Meta-clarification	Examples of asking for clarification about the study procedure
Problem definition	Describes creating an initial or subsequent problem representation that includes potentially useful knowledge elements
Missing information	Represents recognizing and/or searching for clarification about the assignment
Backtracking	Examples of retreating toward an earlier or intermediate state or even to the beginning of the problem
Evaluation	Describes evaluating an explorative or physical operator relative to some task requirement
Operators	Examples of applying mental operators while searching for content and actions
Schema activation	Describe the application of mental structures drawing on past experience
Impasse	Statements refer to a state of mind in which the item writer feels that all options have been exhausted
Solution satisfaction	Describe meeting some desired goals
Constraining	Set limits on the problem solving space
Relaxation	Describe expanding the problem solving options
Decomposition	Describe breaking-up problem-relevant chunks

Study Results

Initially, the frequency of different categories of statements was examined. Separate analyses were completed for each protocol. For each category, the number of statements in that category was computed and the number of statements was expressed as a percent of the total statements. Differences in frequency were examined across categories and also across protocols.

The percentages of different statement categories expressed by each writer are shown in Table 3. For all writers, the greatest percentage of statements was operators. However, the less experienced item writer expressed operators in 31.4% of the statements compared to a mean of 57.1% operator statements for the more experienced writers. Furthermore, the less experienced item writer had more meta-clarification, more missing information, and more impasse statements than the more experienced writers.

Table 3. Statement Category Percentages for Each Writer

	Writer 1*	Writer 2*	Writer 3**
Extraneous	5.6	1.6	19.7
Nonconforming	0.0	1.2	2.1
Meta-clarification	0.0	1.2	7.6
Problem definition	8.0	15.8	11.0
Missing information	7.6	2.2	7.1
Backtracking	0.4	0.9	0.0
Evaluation	6.4	14.9	9.2
Operator	63.9	50.3	31.4
Schema activation	0.0	2.8	1.4
Impasse	0.4	2.8	4.6
Solution satisfaction	4.0	4.0	2.1
Constraining	3.2	1.6	2.1
Relaxation	0.4	0.6	1.8
Decomposition	0.0	0.0	0.0

*Writers 1 and 2 had some experience writing scenario-based tasks.

**Writer 3 had no experience writing scenario-based tasks.

The data show that the experienced writers moved forward in the problem space at a higher frequency than the writer with no experience in scenario-based tasks.

Statements that tend to move the writer forward include evaluation, operator, schema activation, relaxation, decomposition, and solution satisfaction categories. With the exception of problem definition, the remaining categories indicate that the writer may have stalled or moved backward in the problem space. The category of problem

definition is a neutral category that simply indicates that the writer is seeking to define the problem space. Writers 1 and 2 moved forward in the problem space with 75.0% and 73.4% of their statements, respectively. Only 17.3% and 11.6% of their statements, respectively, indicate stalled or backward motion in the problem space. These data are sharply contrasted with the inexperienced writer, who moved forward with 46.1% of the statements and backward with 43.3% of the statements.

In addition to statement frequency, the placement of kinds of categories was examined. The theory predicts a pattern of categories will occur throughout the item writing process. Problem definition statements should occur at the beginning of an item writing phase or activity. The middle of the item writing activity should be dominated by operator statements. The conclusion of the item writing activity should be dominated by evaluation and solution satisfaction statements. This pattern may be repeated throughout the activity until the task is complete.

Category placement data were analyzed using scatter plots that show forward-moving categories placed above problem definition and categories that indicate stalled or backward motion placed below problem definition. These graphs indicate the movement of the writers as they progress through the problem space. The placement of categories for Writer 1, 2, and 3 are respectively shown in Figures 2, 3, and 4.

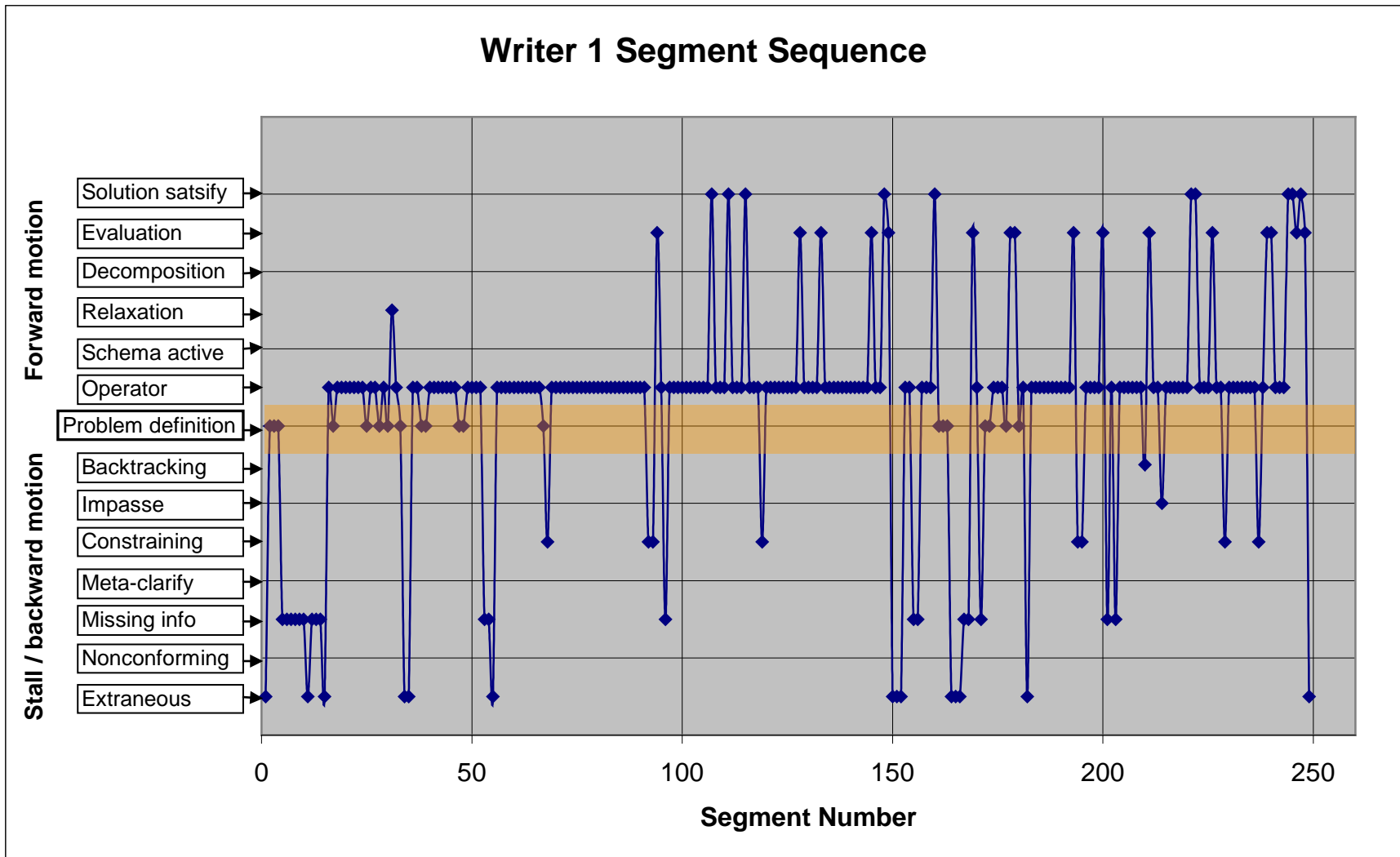


Figure 2. Category placement for Writer 1, indicating movement through the problem space.

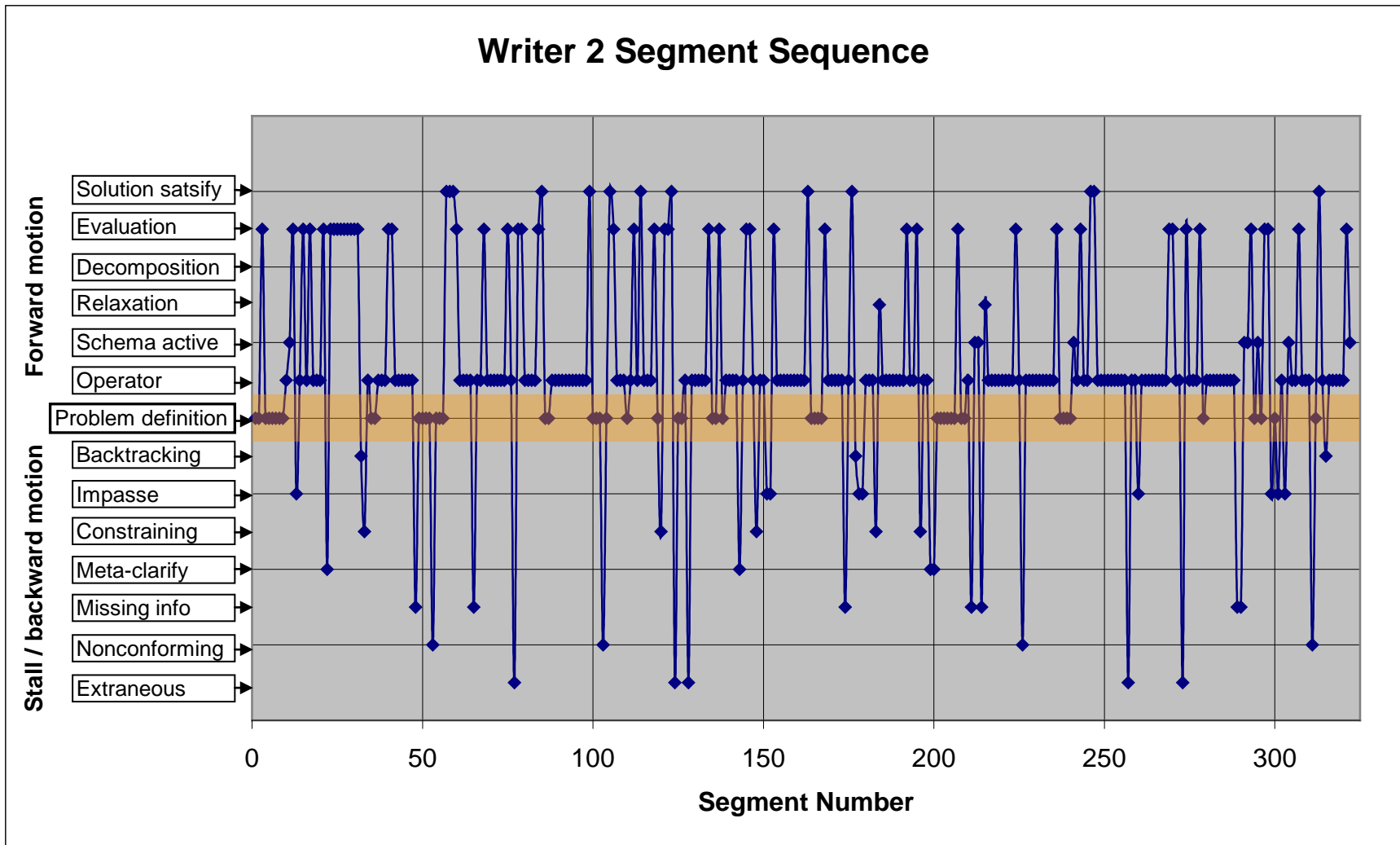


Figure 3. Category placement for Writer 2, indicating movement through the problem space.

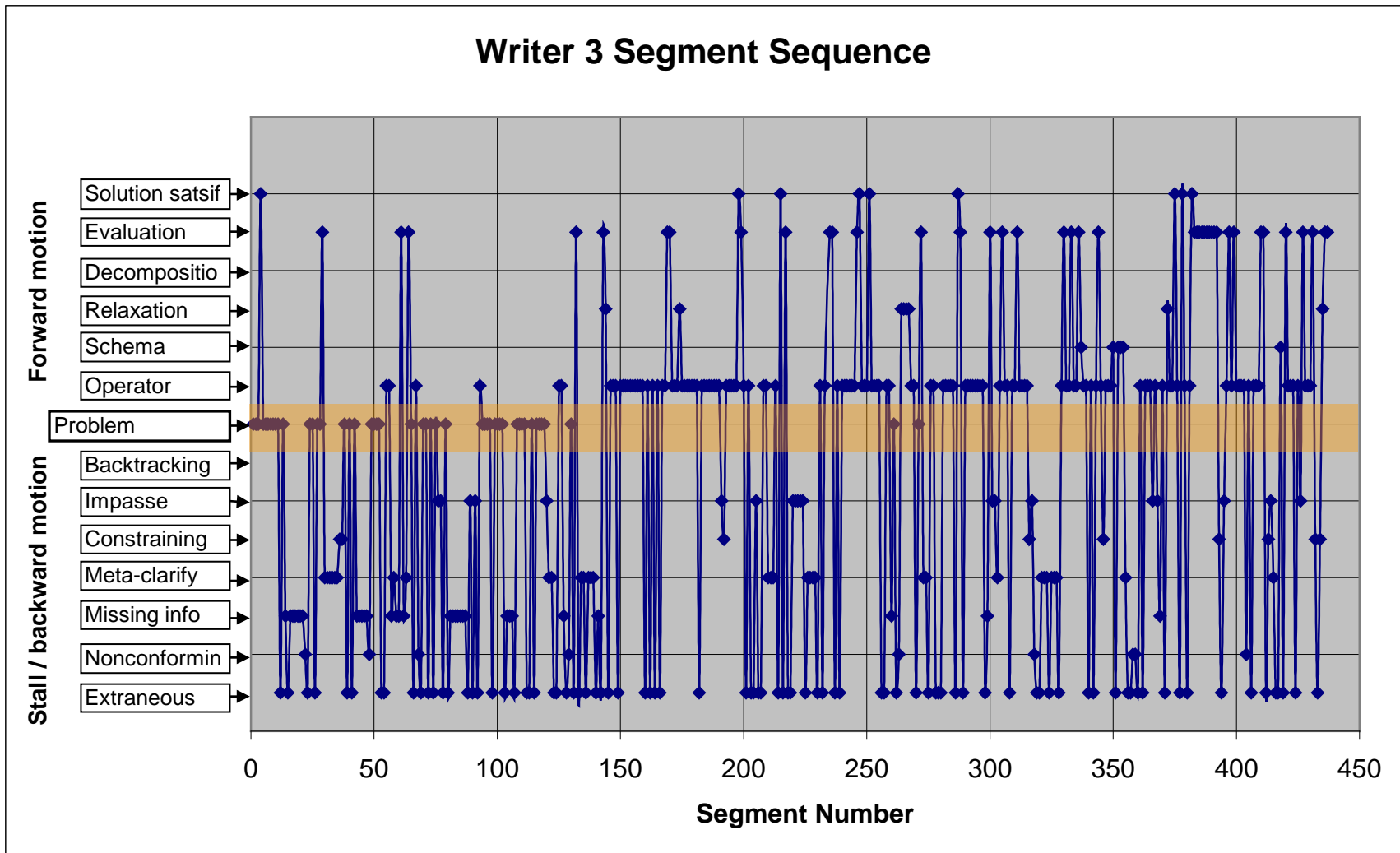


Figure 4. Category placement for Writer 3, indicating movement through the problem space.

The statement analysis of Writer 1 shows category placements that tended to be grouped into well defined stages with generally forward movement through the problem space. A frequent combination of operator, evaluation, and solution satisfaction statements following an initial period of concentrated problem definition can be seen in Figure 2. A similar pattern is shown by Writer 2, whose statement analysis indicates a frequently occurring use of varied combinations of problem definition, operator, evaluation, and solution satisfaction statements (Figure 5). This repeated combination is predicted by the theory.

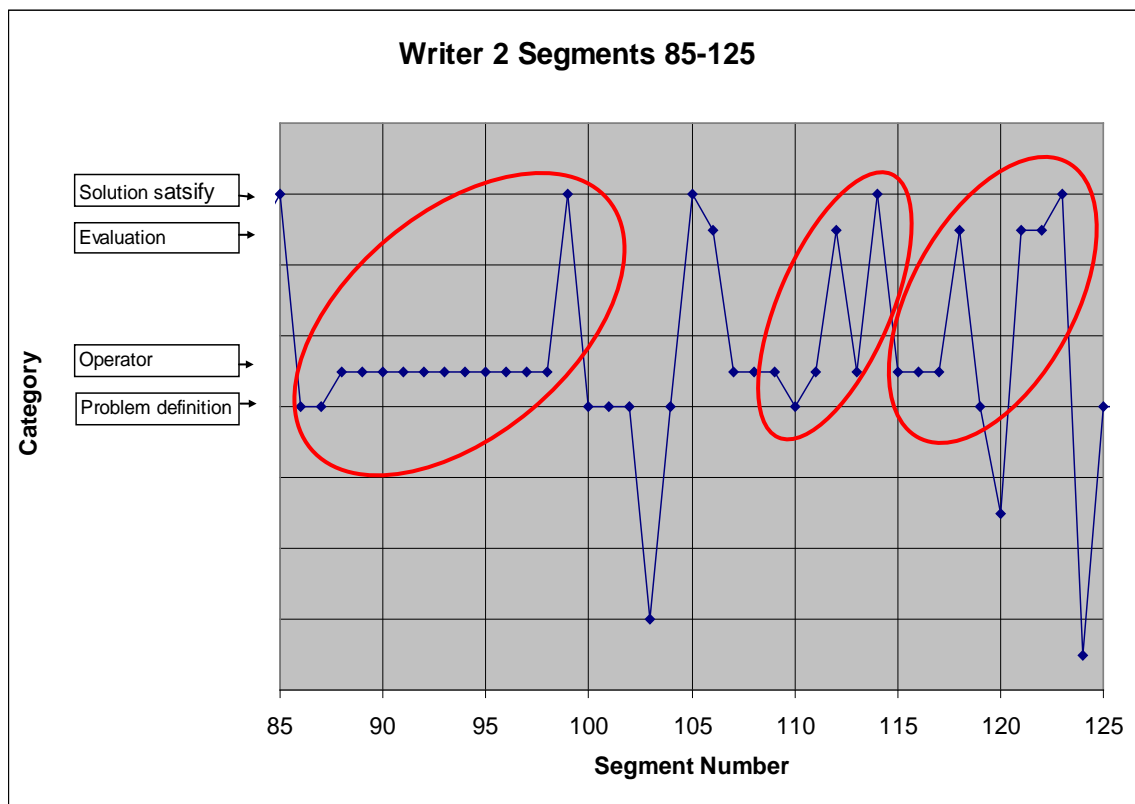


Figure 5. Exemplar theory-predicted statement combinations shown by Writer 2.

A comparison of the placement of categories for Writer 3, the less experienced writer, with the placement of categories for the more experienced writers shows a similar yet less frequent per-statement use of the theory-predicted combination. Additionally, Figure 4 shows an erratic pattern of statement categories when compared to the experienced writers and a relatively large number of statements indicating frequent stalled or backward movement in the problem space.

Conclusion

The goal of the study of item writers is to “bottle” the knowledge and skills acquired by expert writers over years of hard work. The descriptions of the identified conceptual knowledge and skills of expert item writers could be incorporated into item writing workshops in order to better equip new or inexperienced item writers with the tools necessary to produce quality figural response items. Toward that end, this study attempted to understand the patterns of cognition shown by writers with varying degrees of item writing experience in constructing storyboards and writing items for scenario-based tasks.

The results of this study suggest that inexperienced item writers need to be supported in several ways. First, inexperienced item writers may have difficulty accommodating the cognitive load of an item writing assignment. Item writing workshops should explore ways to decrease the cognitive load. Perhaps the assignment could be decomposed. Second, inexperienced item writers have difficulty organizing their response to an assigned task. Item writing workshops could offer new item writers templates and other supports that structure their item writing process. Finally,

inexperienced item writers have a greater demand for information. Item writing workshops should attempt to support new item writers by making readily available information relevant to their assignment, perhaps using pull down menus or other technology.

In addition, this study further developed the theory of item writing first offered by Fulkerson, Mittelholtz, and Nichols (2009). Together, the results of these studies suggest that the development of expertise in item writing is similar to the development of expertise in other domains studied by cognitive scientists.

References

- Ericsson, K.A. & Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge: MIT Press.
- Fulkerson, D., Mittelholtz, D.J., & Nichols, P. D. (April, 2009). *The psychology of writing items: Improving figural response item writing*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Fulkerson, D., Nichols, P., Haynie, K., & Mislevy, R. (2009). *Narrative Structures in the Development of Scenario-Based Science Assessments (Large-Scale Assessment Technical Report 3)*. Menlo Park, CA: SRI International.
- Gentner, D., & Stevens, A.L. (eds.). (1983). *Mental Models*. Hillsdale, NJ: Erlbaum.
- Johnson-Laird, P.N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.
- Minnesota Department of Education (MDE). (2008). *Minnesota Comprehensive Assessment Series II (MCA-II) Test Specifications for Science*. Roseville, MN: Author.

Appendix A
Writing Assignment

DP Study Writing Assignment

Writer Name:

Grade: 8

Type: Investigation

Storyboard: Write a rough 4-scene storyboard (no art necessary) based on the assigned benchmarks. Use the storyboard scene template.

Item: Write one rough multiple-choice item aligned to one of the assigned benchmarks in the context of one of your storyboard scenes. Use the item template.

Assigned Benchmarks: Please select one of the following groups of benchmarks as your storyboard/item writing assignment for the study. You may add additional benchmarks to your assignment if desired.

Physical Science

- Choose 1 of the following: 6.I.B.1 & 7.I.B.1 or 6.I.B.4
- 6.II.A.3
- 6.II.B.1 & 6.II.B.2

Earth Science

- Choose 1 of the following: 6.I.B.1 & 7.I.B.1 or 6.I.B.4
- 8.III.A.4
- 8.III.A.6

Life Science

- Choose 1 of the following: 6.I.B.1 & 7.I.B.1 or 6.I.B.4
- 7.IV.A.2 & 7.IV.A.3
- 7.IV.A.6

Design Pattern: Yes No

If yes: Observational Investigation Experimental Investigation

Appendix B

Instructions to Experimenter and Participant

Participant Name:

Time:

Date:

Instructions to experimenter

Begin the session.

Instructions to Participants

Thank you for helping us pilot the design patterns. In this pilot study, you will be asked to think aloud as you create four scenes and a multiple-choice item aligned to the scene and the selected benchmarks. You will be presented with the instructions in just a moment. But first, we ask that you think aloud during this study:

Do not narrate

Do not explain

Do not edit

Do not worry about word choice

Just say what you are thinking

Whatever is going through your mind

Practice

Let's practice with a multi-column subtraction item. Feel free to mark on any of the materials. Start to think aloud as you are handed the item. Think aloud as you solve the item. Do not pause, but continue to talk. Say whatever you are thinking.

806

- 344

Good. Remember, do not pause but continue to think aloud as you solve the item. Let's try another multi-column subtraction item.

322

- 167

Four Scenes

Now that you have completed the practice task, you will be given the instructions and some materials. Begin to think aloud as soon as you start to read the instructions. Read aloud the instructions. Read aloud any other materials as you read them and use them. Finally, think aloud as you create the scenes aligned to the selected benchmarks.

- Write a 4-scene rough draft storyboard supporting assigned benchmarks
- Use the design pattern if required
- “Think aloud” while writing

Instructions to experimenter

Initially, handout the following:

- Study Writing Assignment
- Storyboard scene template
- If they did not bring a copy of the benchmarks, then provide a copy
- If they did not bring a copy of the design pattern (if necessary), then provide a copy

- They may bring a copy of the Training PowerPoint. DO NOT provide a copy.

DO NOT ALLOW OTHER MATERIAL, SUCH AS TEXTBOOK.

Please make a note if they bring the PowerPoint.

Allow participants 45 minutes to complete the four scenes. After 45 minutes, move on to the multiple-choice item.

Prompt the participant if they fail to think aloud for longer than 5 seconds.

Instructions to experimenter

After the participant has completed the storyboard scenes, ask them to select a scene and a benchmark. Then write a rough draft multiple-choice item aligned to the scene and benchmark.

- Write one rough draft MC item aligned to one storyboard scene and one assigned benchmark
- Use the design pattern if required
- “Think aloud” while writing

If the participant jumps to the item, do not stop and repeat instruction. Provide them with the template and remind them to think aloud.

Instructions to Participants

Now that you have completed the four scenes, please write one rough draft of a multiple-choice item. Please review the instructions. Begin to think aloud immediately. Read aloud the instructions. Read aloud any other materials as you read them and use them. Finally, think aloud as you create the multiple-choice item.

Be sure the participant marks or indicates the benchmark to which they are writing.

Instructions to experimenter

End the session.

Appendix C

Example Storyboard Scene Template

Scene #: _____

Scene text:

Detailed scene description:

Content or art references:

Assessed benchmarks:

Appendix D

Example Multiple-Choice Item Template

Item #: _____

Item stem text:

Answer options text:

A.

B.

C.

D.

Rationales:

A.

B.

C.

D.