An Evaluation of *Simulations in Developmental Disabilities (SIDD)*: Instructional Software that Provides Practice in Behavioral Assessment and Treatment Decisions

Marcie N. Desrochers
Tiffany Clemmons
Mary Grady
Brandi Justice

ABSTRACT. *Simulations in Developmental Disabilities (SIDD)* is a multimedia computer program designed to provide undergraduate psychology students with practice in making assessment and treatment decisions. Eighteen undergraduate psychology students participated in an experiment to test the instructional effectiveness of *SIDD*. Post-test scores were significantly higher in the experimental group who received training with *SIDD* than in a control group who did not receive training. The students also rated the software positively. Future strategies to further evaluate the software are discussed. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-342-9678. E-mail address: <getinfo@haworthpressinc.com> Website: <http://www.HaworthPress.com>*

KEYWORDS. Computer-aided instruction, simulation, functional assessment, behavioral treatment, software evaluation

---

Marcie N. Desrochers, PhD, teaches psychology at the College of Charleston. Tiffany Clemmons and Mary Grady both have their BS degrees in psychology from the College and Brandi Justice is currently working on her BS.

Address correspondence to: Marcie Desrochers, College of Charleston, Department of Psychology, 66 George St. Charleston, SC 29424 (E-mail: desrochersm@cofc.edu).

Support for this research was provided by a NIH AREA grant.

Traditionally, clinical skills have been taught through textbook, lecture, and/or practicum formats. Supplementing these modes of instruction with the use of a simulation technique may enhance the student's acquisition of clinical skills (e.g., Holsbrink-Engels, 1997). There are many advantages to using computer simulations as an instructional tool, including that: (a) students are able to experience events without risk of harm or damage to themselves or others; (b) unique situations can be explored; (c) it is a time saving device that can rapidly display complex phenomena; (d) effective teaching strategies can be easily programmed (e.g., immediate feedback, learner control, and individualized student participation); and (e) students' responses may be readily recorded and compared (Flynn, 1990; Jelovsek, Catanzarite, Price, & Stull, 1989; Jones & Keith, 1983; Kinzle, 1990; Miller, 1984). Moreover, incorporating a multimedia component into problem-based learning may enhance its effectiveness by presenting realistic and relevant stimuli to the learner (Hoffman & Ritchie, 1997; Lalley, 1998). Although the possible benefits are many, whether they apply to a particular software program must be empirically determined.

Formative and summative evaluations can be conducted to evaluate instructional software (Mark & Greer, 1993). Generally, formative evaluations are performed during the early stages of program development to generate information about the operation of the software and determine whether it functions effectively from the user's standpoint. Formative evaluations may involve gathering reviewers' ratings of the usefulness and effectiveness of that software (e.g., Desrochers & Hile, 1993) or conducting studies designed to determine the content validity of the software (Desrochers, Hile, & Williams-Moseley, 1997).

Formative evaluations may not always be the best approach to evaluate software, however. Some research suggests that reviewers' evaluations may not accurately assess the software product and are influenced by contextual factors like previous software used (Jones & McCormac, 1992). Other research suggests that subjective evaluations do not always agree with objective performance measures of software outcome (Stark, Gruber, Renkle, & Mandl, 1998; Zahner, Reiser, Dick, & Gill, 1992).

An alternative or supplement to formative evaluations is summative evaluation, which refers to using more formal experimental procedures designed to answer questions such as, whether the software is effective in achieving its intended purpose, or to what degree do students acquire the target material. Although this has not always been

the case in the history of computer instruction, more and more software is currently receiving summative evaluation (Fletcher-Flinn & Gravatt, 1995; Mark & Greer, 1993).

For example, Holsbrink-Engels (1997) conducted an experiment to evaluate computer-based role-playing. Significantly greater change scores (post-test minus pre-test scores) for students' knowledge and classification tests occurred in the experimental group, but not the control group. Based on subjective ratings, students in the experimental group were satisfied with the manageability of technical aspects of the system, usefulness of the program for skills training, and effectiveness of manual, lesson, and feedback.

In another study, Overbaugh (1995) found a significant change in education majors' pre-test compared to post-test scores when evaluating the effectiveness of interactive-video computer aided instruction (CAI) designed to teach classroom management skills. Stark et al. (1998) compared multiple versus uniform learning situations and guided versus unguided problem solving with 60 economics students. These researchers found that the guided multiple problems produced better student performance measures compared to outcomes in the other conditions.

This study was conducted to empirically evaluate the instructional effectiveness of *Simulations in Developmental Disabilities (SIDD)*, a multimedia software designed to provide undergraduate psychology students with practice in making behavioral assessment and treatment decisions. Pre- and post-test scores for an experimental group who received instruction with *SIDD* were compared to a control group who did not use the software. Participants' subjective evaluations of the software regarding the usefulness and effectiveness of the software in achieving its instructional objectives were also collected.

METHOD

Participants

Eighteen college students from a small liberal arts college participated in the study for extra course credit. All participants were upper level psychology majors who were enrolled in at least one of Applied Behavior Analysis, Research Methods, or Systems and History course. Descriptions of participants are presented in Table 1.

TABLE 1. Subject Characteristics

		Experimental Group	Control Group
Mean age:		20.44	23.11
Gender:	Female	5	6
	Male	4	3
Race:	Caucasian	5	8
	African-American	3	0
	Asian-American	0	1
Grade point average:	4.0-3.6	0	5
	3.5-3.1	3	0
	3.0-2.6	3	2
	2.5-2.0	3	2
Courses currently or previously enrolled in:	Conditioning & Learning	7	8
	Research Methods	8	7
	Applied Behavior Analysis	3	5
Experience with persons with mental retardation		1	1

Materials and Apparatus

The study was conducted in two computer classrooms. One room contained 30 Zenith Data System Z-Station Pentium 100 MHz desktop computers with 32 MB RAM, CDR-SIG 4X-speed CD-ROM, and Windows 95. The other room contained 12 Dell Pentium 100 MHz desktop computers with 32 MB RAM, 24X CD-ROM, and Windows 95.

Software. SIDD is a multi-media instructional software designed to provide students with practice in: (a) operationally defining a simulated client's target behavior; (b) selecting relevant assessments; (c) interpreting graphed data of the client's behavior; (d) developing a functional hypothesis concerning the cause of the client's problem behavior; and (e) formulating an effective treatment plan. During the simulation, the student assumes the role of a clinician and proceeds through referral, problem definition, assessment, functional hypothesis, and treatment stages with a particular client case. Video clip excerpts of the simulated client can be viewed during assessment and follow-up (see Desrochers & Hile, 1993 for a detailed description of this software). ''User friendly'' features of the software include help options, notepad, screen instructions and hypertext-linked terminology. The software has an accompanying tutorial that reviews the behavioral

terminology found in *SIDD* by presenting multiple choice and fill-in-the-blank questions to the user, and providing explanations for incorrect responses.

Simulated Client Tests. Tests were used to evaluate the participant's behavioral assessment and treatment knowledge and were administered before (Test 1), immediately after (Test 2), and 2 weeks following use of *SIDD* (Test 3). Each test included general referral information for one simulated client, graphed functional analysis information (see Iwata, Dorsey, Slifer, Bauman, & Richman, 1994), and a series of questions concerning the client. Participants were requested to: (a) define the client's problem behavior; (b) indicate what assessments should be conducted; (c) analyze graphed data and identify a cause of the client's problem behavior; and (d) provide treatment recommendations. A maximum test score of 15 could be earned. Each test administered to the participant consisted of a different simulated client. Simulated clients could differ in terms of characteristics (e.g., age, gender), problem behavior (self-injurious behavior, aggression), cause of problem behavior (e.g., attention, escape from work, self-stimulation), location of the problem (school, home, vocation setting), and referral agent (parent, teacher). The client cases comprising the tests were not used during training sessions.

Procedure

The study was conducted twice, with one group of 12 students (Study 1) and another group of 6 students (Study 2). Since no systematic differences in participant characteristics or results occurred between the two studies, the data was merged.

Participants were randomly assigned to experimental and control groups. An exception occurred in Study 1 where two participants opted to participate in the experimental group due to conflicting time commitments. Also, unequal assignment of participants to conditions occurred in both studies due to participants failing to show up for sessions (i.e., three participants failed to attend the first session in Study 1 and one in Study 2).

The same procedure was employed with both studies. During the first session, a general introduction to the nature of the study was presented to participants and then both experimental and control groups received Test 1. Also during the first session, only the experimental group received training with the *SIDD* tutorial software. Dur-

ing the next four sessions that were approximately one hour in duration and held on separate days, only the experimental group received training with four client cases using the *SIDD* software. Following completion of the *SIDD* training by the experimental group, both experimental and control groups received Test 2. Test 2 was the same as Test 1 except that a different simulated client was used (see Simulated Client Tests for a description). Past research suggests that difficulty level of simulations is an important variable determining student performance (e.g., Scheuneman, Fan, & Clyman, 1998). To control for difficulty level of client cases, order of client cases trained in *SIDD* and simulated client tests were counterbalanced within groups for Test 1 and Test 2. A follow up test (Test 3), with a different simulated client than those used for Test 1 and Test 2, was administered 2 weeks following Test 2 with the experimental subjects. Following Test 2, the participants in the control group received training with the *SIDD* tutorial and four *SIDD* client cases that were not used during testing.

Using answers derived from an independent expert, participants' tests were scored by a research assistant and compared across experimental and control groups. To lessen experimenter bias, information concerning the identity, condition, and number of each test was concealed from the scorer. The first author provided training to the scorers and conducted random spot checks of the scoring system. An analysis of covariance (ANCOVA) with Test 1 scores as covariate was applied to determine whether significant differences existed for Test 2 scores between conditions.

Qualitative Evaluation. Subjective evaluations of the software were collected after the experimental and control group participants had completed the final test following use of the *SIDD* software. This questionnaire asked participants to rate, along a 5-point Likert-type scale (1 = *Extremely Useful/Effective* to 4 = *Not At All Useful/Effective* with 5 = *Undecided*), the software in terms of: (a) mechanics of the program (i.e., How easy or difficult the program is to use? Were the instructions and help option useful?) and; (b) perceived instructional effectiveness of the program (e.g., How effectively does SIDD teach conducting functional assessments? How useful is the software for teaching behavioral assessment skills? Compared to traditional instructional formats (e.g., textbook, lecture) how useful is the software for teaching behavioral assessment skills?).

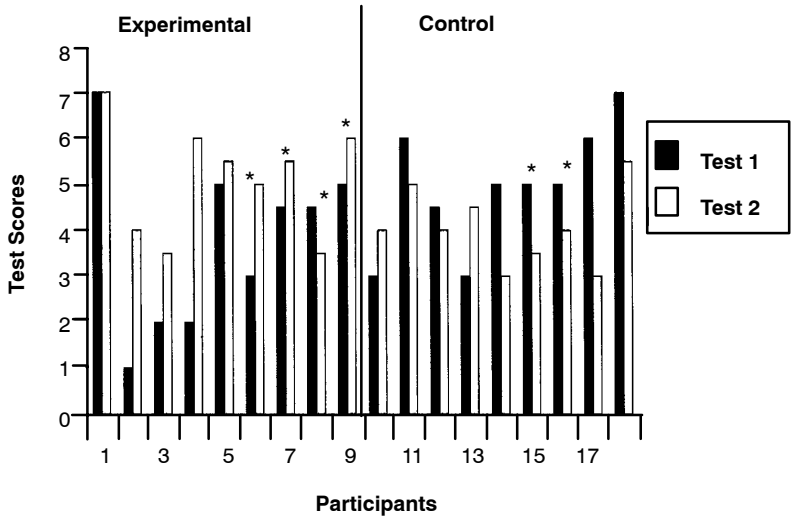
Procedural Reliability. A second scorer, who was also unaware of the participants' identities, test number, and condition, independently scored 25% of the tests. Interobserver reliability of scoring was calculated by examining the number of occurrences in agreements in scoring of each open-ended answer to a test question over the number of agreements and disagreements and multiplying by 100. The average procedural reliability score was 84.9%, ranging from 58.3% to 100% for a particular test.

A more conservative measure of reliability, Cohen's kappa, was also applied to the data. A kappa score of $k = 0.49$ was found, which is statistically significant $z(0.18) = 2.78$, $p < .05$ or indicative of a reliable outcome.

RESULTS

Using the *SIDD* software seemed to improve participants' test scores. As seen in Figure 1, Test 2 scores are generally higher than Test 1 scores for participants in the experimental group compared to those in the control group (see Figure 1). The experimental group

FIGURE 1. Test 1 and Test 2 Scores for Each Participant in the Experimental and Control Groups. Asterisks Indicate Participants from Study 2.



mean Test 1 score was 3.78 (SD = 1.91) and mean Test 2 score was 5.11 (SD = 1.22) while for the control group the mean Test 1 score was 4.94 (SD = 1.33) and mean Test 2 score was 4.06 (SD = 0.84). No significant differences between groups for Test 1 mean scores were found as determined through an independent-groups 2-tailed *t* test ($t(16) = -1.51, p > .15$). Although Test 2 scores were on average lower than Test 1 scores for the control group, this difference was also not significant (paired sample *t* test ($t(8) = 1.89$, $p > .10$). An ANCOVA, with covariate of Test 1 scores and dependent measure of Test 2 scores, was applied and a significant effect of using *SIDD* was found in the experimental group compared to the control group, $F(1,15) = 8.71$, $p < 0.01$. Follow up Test 3 scores remained high with a mean of 5.72 (SD = 1.28) found for participants in the experimental group.

Subjective evaluations were collected from participants in both conditions following use of the *SIDD* software. In terms of functioning of the software, most participants rated the software *very easy* to use (see Figure 2). In terms of content, most participants rated *SIDD* as *extremely* or *very effective* in teaching functional assessment (see Figure 3). Participants found the software to be *extremely useful* for the application of behavior principles (see Figure 4) and *extremely useful* compared to traditional instructional formats (see Figure 5).

DISCUSSION

The results suggest that, in general, the *SIDD* software was effective in improving the assessment and treatment decisions made by the undergraduate psychology students who participated in this study. Furthermore, the students perceived the software to be useful and effective in meeting the assigned instructional objectives. In contrast to previous research (Jones & McCormac, 1992; Stark et al., 1998), the objective and subjective outcomes agreed in this particular case.

Several aspects of this study need to be considered. First, why did test scores for any particular participant not reach the maximal number of possible points? Students participated in this study for extra credit and no consequences for a certain level of performance on tests were delivered. It is possible that participant motivation might be an issue in this study. Perhaps higher test scores might have occurred if participants were paid for participating in the study or if the software was tested as part of a course requirement. Morrison, Ross, Gopalakrish-

FIGURE 2. Participants' Evaluations of Ease of Use of the SIDD Software.

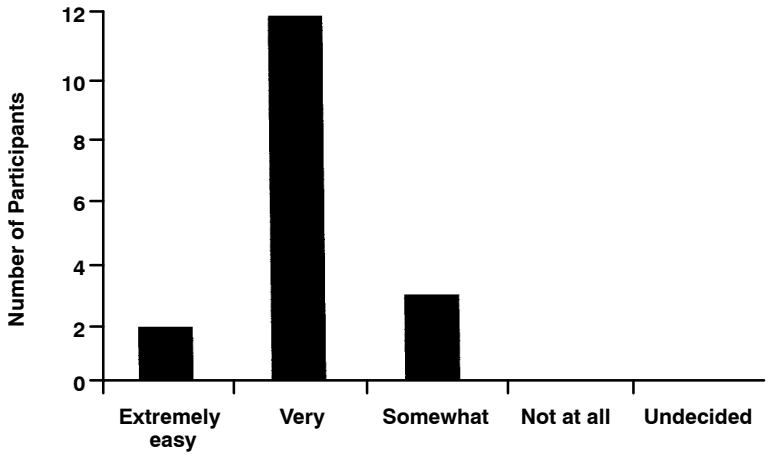


FIGURE 3. Participants' Evaluations of the Effectiveness of SIDD in Teaching Functional Assessment Skills.

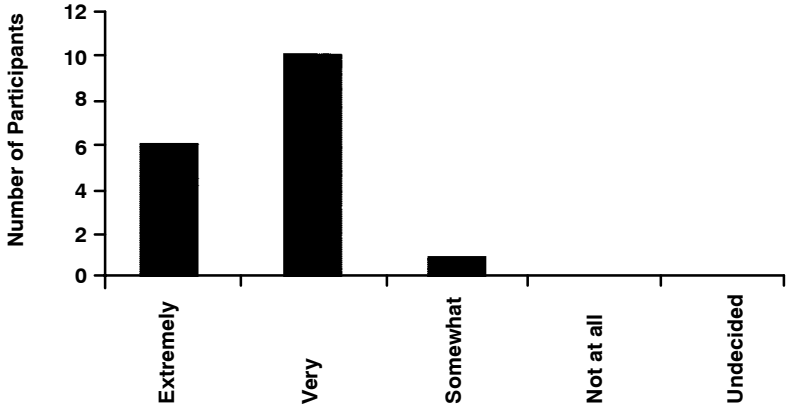


FIGURE 4. Participants' Evaluations of Degree of Practice in Application of Behavior Principles and Procedures Provided by SIDD.

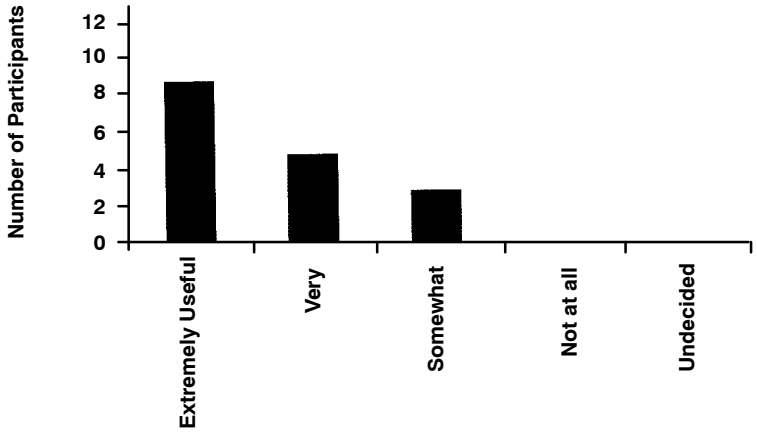
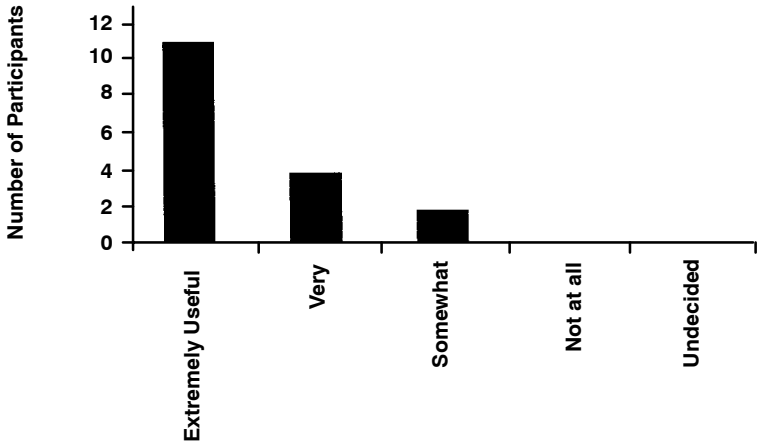


FIGURE 5. Participants' Evaluations of SIDD Compared to Traditional Instructional Formats.



nan, and Casey (1995) examined three types of feedback strategies, incentives, and types of answers with 246 introductory teacher education students. They found that the students who used the software as part of their course requirement performed better than those who earned extra credit for their participation in the study.

A second issue is that Test 1 scores differed slightly (although nonsignificantly) between groups, even though participants were randomly assigned to groups (with the exception of two students). It is possible that this difference occurred as a result of several more students who were enrolled in an Applied Behavior Analysis course comprising the control group compared to the experimental group. Research examining the effectiveness of *SIDD* with a restricted sample of students who have comparable levels of background in applied behavior analysis (e.g., students in an applied behavior analysis class for which the software was designed) might provide a clearer test of the usefulness of the software. On the other hand, examining the performance of students with varying levels of knowledge of behavioral psychology provides a measure of the external validity of the software.

There was also a slight, although nonsignificant, decrease in Test 2 scores compared to Test 1 scores for 7 of the 9 participants in the control group despite counterbalancing content of the tests within groups. This difference in mean scores for the control group may have occurred simply due to chance. On the other hand, since the differences occurred with most of the control group participants, it may reflect a real but small difference. If so, lower scores on the second testing may have occurred due to the lack of feedback given to students following Test 1, in other words, it may represent an ''extinction'' effect (see Martin & Pear, 1999). It is also possible that since participants were not blind as to which condition they were in, that subject bias may have been present. Further investigation of this issue is needed with a larger sample of participants, firstly, to determine whether differences in scores on repeated testing are real and, secondly, to identify the conditions under which it occurs.

Future investigations might also entail comparing *SIDD* to more traditional methods of instruction (e.g., reading case studies). Additionally, examining *SIDD*'s usefulness as an assessment tool for participants with differing levels of behavioral knowledge (e.g., undergraduate versus graduate students versus practitioners) could be conducted. Mark and Greer (1993) consider this type of sensitivity analysis to be either formative or summative depending on how the results are used. If such a study is implemented to help with program design, then formative evaluation is involved. If the study is used to evaluate the intended purpose of the software as an assessment tool, then it would be considered summative evaluation.

Software evaluation, whether formative or summative, is a necessary step in the development of effective instructional tools. Much useful information can be yielded from this process to create software that will ultimately be of benefit to the learner.

REFERENCES

Desrochers, M.N., & Hile, M.G. (1993). SIDDS: Simulations in developmental disabilities. *Behavior*, *Research Methods*, *Instruments*, *& Computers*, *25*(2), 308-313.

Desrochers, M.N., Hile, M.G., & Williams-Moseley, T. (1997). A survey of functional assessment procedures used with individuals with severe problem behaviors and mental retardation. *American Journal on Mental Retardation*, *101*, 535-546.

Fletcher-Flinn, C.M., & Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): A meta-analysis. *Journal of Education Computing Research*, *12*, 219-242.

Flynn, J.P. (1990). Using the computer to teach and learn social policy: A report from the classroom and the field. *Computers in Human Services*, *7*, 199-209.

Hoffman, B., & Ritchie, D. (1997). Using multimedia to overcome the problems with problem based learning. *Instructional Science*, *25*, 97-115.

Holsbrink-Engels, G.A. (1997). Computer-based role-playing for interpersonal skills training. *Simulation & Gaming*, *28*, 164-180.

Iwata, B.A., Dorsey, M., Slifer, K., Bauman, K., & Richman, G. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, *27*, 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities*, *2*, 3-20, 1982)

Jelovsek, F.R., Catanzarite, V.A., Price, R.D., & Stull, R.E. (1989). Application of teaching and learning principles to computer-aided instruction. *CAI in Medicine*, *6*, 267-273.

Jones, G.L., & Keith, K.D. (1983). Computer clinical simulations in health sciences. *Journal of Computer-Based Instruction*, *9*, 108-114.

Jones, B.T., & McCormac, K. (1992). Empirical evidence shows that measuring users' opinions is not a satisfactory way of evaluating computer-assisted learning in nurse education. *International Journal of Nursing Studies*, *29*, 411-425.

Kinzle, M.B. (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation, and continuing motivation. *Educational Technology Research & Development*, *38*, 1-21.

Lalley, J.P. (1998). Comparison of text and video as forms of feedback during computer assisted learning. *Journal of Educational Computing Research*, *18*, 323-338.

Mark, M.A., & Greer, J.E. (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, *4*, 129-153.

Martin, G., & Pear, J. (1999). *Behavior modification: What it is and how to do it* (6th Ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Miller, M.D. (1984). The use of simulations in training programs: A review. *Educational Technology*, *24*, 39-42.

Morrison, G.R., Ross, S.M., Gopalakrishnan, M., & Casey, J. (1995). The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology*, *20*, 32-50.

Overbaugh, R.C. (1995). The efficacy of interactive video for teaching basic classroom management skills to pre-service teachers. *Computers in Human Behavior*, *11*, 511-527.

Scheuneman, J.D., Van Fan, Y., & Clyman, S.G. (1998). An investigation of the difficulty of computer-based case simulations. *Medical Education*, *32*, 150-158.

Stark, R., Gruber, H., Renkl, A., & Mandl, H. (1998). Instructional effects in complex learning: Do objective and subjective learning outcomes converge? *Learning and Instruction*, *8*, 117-129.

Zahner, J.E., Reiser, R.A., Dick, W., & Gill., B. (1992). Evaluating instructional software: A simplified model. *ETR&D*, *40*, 55-62.

RECEIVED: 9/07/99
ACCEPTED: 28/10/99