Automated Support for Learning in Simulation: Intelligent Tutoring of Shiphandling

Stanley Peters, Elizabeth Bratt
Stanford University, CSLI
Stanford, California
peters@stanford.edu, ebratt@stanford.edu

Susan Kirschenbaum
Naval Undersea Warfare Center
Newport, Rhode Island
susan.kirschenbaum@navy.mil

ABSTRACT

Immersive VR simulations provide realistic experience in skills that are expensive or dangerous to learn principally through real world practice, for instance conning large ships. Although the cost of simulators can be amortized over many training sessions, the ongoing cost for human instructors is a serious impediment to optimal utilization of simulations. An artificially intelligent tutoring system (ITS) capable of performing some of an experienced human instructor's role may, without diminishing student proficiency, enable a reduction in dollar and manpower costs by helping students overcome common and predictable problems during simulation-based training. ONR supported an R&D effort to develop and test such an ITS for the Conning Officer Virtual Environment (COVE) shiphandling simulator used to train Naval Surface Warfare Officers (SWO). COVE ITS detects both process (observation of visual cues) and performance (correct orders at the correct time) as conning officers conduct a briefed evolution in COVE. It measures the student's proficiency and detects problems that may arise. Modeled on instructor behavior, the ITS can give pointers before problems become unmanageable, but not before the student has a chance to see the effects of an error. This system is being evaluated at the Surface Warfare Officers School (SWOS). A study to assess proficiency gains of students taught by COVE ITS with oversight by an instructor, compared to proficiency gains of a control group who received the usual one-on-one instructor training used two runs, one instructed and a second uninstructed test run. Data analyzed to date show no differences in performance between the two groups during the test run, either on instructor scoring or on behavioral measures. Additional evaluation will test the effectiveness of the ITS when one instructor supervises two or three COVE stations. Plans for further development include expansion to different levels of students and to a practice-only mode.

ABOUT THE AUTHORS

Stanley Peters is professor emeritus of linguistics at Stanford University and director of its Center for the Study of Language and Information. He has a research focus on interactive spoken language technology, including artificially intelligent tutoring systems. Peters is a Fellow of the Linguistic Society of America.

Elizabeth Owen Bratt is a Senior Research Engineer in the Computational Semantics Laboratory of the Center for Study of Language and Information (CSLI) at Stanford University. She received her Ph.D. in Linguistics from Stanford University. Her research at CSLI has focused on how natural language interfaces can support learning in intelligent tutoring systems. Previously, she was a Research Linguist at SRI International, working in concept-to-speech generation in dialogue systems and development methods for spoken language understanding in domains with limited data.

Susan Kirschenbaum is an Engineering Psychologist in the Combat Systems Department of the Naval Undersea Warfare Center Division, Newport, Rhode Island, and the Human Systems Integration (HSI) Lead for NUWCDIVNPT. Dr. Kirschenbaum received her Ph.D. in experimental psychology from the University of Rhode Island. She has continuing interests in expertise and in the effects of information variables on decision making. She is a Fellow of the American Psychological Association, was the 1996 recipient of the NUWC Independent Research Award and has been a Visiting Scientist in both the US and Australia.

Automated Support for Learning in Simulation: Intelligent Tutoring of Shiphandling

Stanley Peters, Elizabeth Bratt
Stanford University, CSLI
Stanford, California
peters@stanford.edu, ebratt@stanford.edu

Susan Kirschenbaum
Naval Undersea Warfare Center
Newport, Rhode Island
susan.kirschenbaum@navy.mil

INTRODUCTION

We describe a novel approach to integrating an automated instructor with an advanced simulator to reduce the need for human instructors to guide students during training, preserving the level of proficiency students attain.

Complex immersive high-fidelity simulators are now widely used to afford students realistic experience in acquiring skills that are expensive or dangerous to learn solely through real world practice in tasks like operating nuclear power plants, piloting jet airplanes, and conning large ships. Effective use of these simulators in training is currently instructor-intensive, because students are often slow or even unable without instructor support to comprehend complex interactions among multiple factors that result in the simulated response their actions produce, or to recognize the full range of information sources they should be monitoring as relevant to recent or impending actions.

An important motivation underlying this research is the ongoing cost for human instructors to guide students and prevent them from "practicing mistakes," which is a serious impediment to optimal utilization of simulation in training, even though the cost of the simulators themselves can be amortized over many training sessions. Our hypothesis is that an artificially intelligent tutoring system (ITS) capable of performing some of an experienced human instructor's role can, without diminishing student proficiency, enable a reduction in dollar and manpower costs by helping students overcome common and predictable problems during simulation-based training.

The U.S. Navy confronts this cost dilemma in training more than a thousand officers each year with the Conning Officer Virtual Environment (COVE) shiphandling simulator at its Surface Warfare Officers School (SWOS) in Newport, Rhode Island. The Office of Naval Research (ONR) therefore supported research to develop an ITS for COVE and test it at SWOS. This paper describes the prototype COVE ITS that was developed and presents preliminary results from its evaluation at SWOS.

The remainder of the paper proceeds as follows. The next section describes an analysis of a conning officer's tasks and our approach to artificially intelligent tutoring of the shiphandling skills they involve, as well as integration of the ITS with the COVE simulator. Section 2 describes an ongoing trial use of this ITS at SWOS, begun in the first quarter of 2011, and discusses findings from its use in the first three courses. Finally, section 3 summarizes our conclusions to date and our plans for further research.

SHIPHANDLING KNOWLEDGE AND SKILLS

Conning officers typically drive large ships by ordering (i) a helmsman to steer in a certain way, (ii) the lee helm to obtain specific propulsion from the ship's engine(s), and (iii) sometimes also others to apply forces with tugs or with lines or cables attached to buoys or piers or anchors. Such orders are issued to move the ship along a planned trajectory, and the choice of orders should result from (a) detailed observations of the ship's state in relation to its surrounding environment and intended trajectory, together with (b) knowledge of how the ship will respond to particular combinations of forces under the conning officer's command.

Expertise in Shiphandling

The ship's response is causally determined by hydrodynamic and other physical forces, and officers receive lectures on the practical application of relevant principles. However, the physical interactions involved are too complex, and dependent on characteristics of specific ship types, for humans to work out from first physical principles in real time what the effect would be of combinations of rudder, engine, and tug orders. Therefore, it is essential for conning officers to acquire type-specific practical knowledge of interactions among forces on their ship and combinations of orders that will yield desired effects under various conditions. Helping conning officers acquire, refine, and master this knowledge is an important goal of training in shiphandling simulators.

Our research team observed and videotaped SWOS instructors teaching junior conning officers, Navy Ensigns, in the COVE simulator during the practicum sessions of basic shiphandling courses. By analyzing these data we identified how these instructors help students learn their ship's response to rudder, engine, and tug orders—through a combination of questions, hints, directing visual attention, and direct explanations. We also identified instructor strategies for deciding when to allow students to persist in an erroneous order and when instead to intervene in one of these ways. We further noted when instructors would ask a question or offer encouragement in connection with students' correct orders.

This analysis informed the design of COVE ITS, in particular the means by which it would detect moments when to tutor students. In order to discriminate between correct and incorrect student actions, the prototype COVE ITS incorporates expert knowledge about how required ship responses can be produced. Once the ITS has determined the correctness of a student's orders (or omission of orders), it follows an instructional strategy to decide whether to allow the student to continue without comment, or to give the student corrective or reinforcing oral feedback at that moment. The expert knowledge of shiphandling that COVE ITS incorporates was distilled from a combination of standard reference works (including Barber, 2005; Crenshaw, 1975; Noel, 1989; Stavridis & Girrier, 2007) as well as advice from and observation of subject matter experts (SMEs), especially master mariners employed by SWOS to teach shiphandling to senior Naval officers in advanced courses.

Student conning officers must also learn how to determine by observing the ship's state and relationship to its environment what alterations of ship motion if any are necessary for the ship to continue following the planned trajectory or, if it has deviated from that, to return to a satisfactory trajectory. The conning officer must visually observe the environment along with displays on the bridge, obtain reports from observers at other positions aboard ship, and integrate the information thereby obtained into a 'mental picture' of the ship's situation with respect to its environment and its planned trajectory. Some useful visual cues are subtle, and the conning officer must look for them in the right place at the appropriate time. Helping conning officers acquire, refine, and master these skills is another important goal of training in shiphandling simulators, and one of the motivations for employing virtual reality. One reason that critical visual cues are so subtle is the enormous inertia of large ships, which leads to lengthy latencies in responses to some orders. A conning officer who cannot discern the subtle early indications of his order having its desired effect is likely to apply forces that are too large or leave forces in effect too long, only to discover after a while that he has overdone by quite a large margin what he meant to do.

To aid students in internalizing a good cognitive model of shiphandling, instructors and the ITS can employ a behavioral model of when to look where for essential timely information about the progress of a shiphandling evolution. We distilled such a model from a combination of SME advice, including think-aloud shiphandling exercises by master mariners, together with analysis of videotaped lessons in SWOS's basic shiphandling courses. For beginning conning officers, a great deal of instructional effort is devoted to teaching observing skills, so a shiphandling ITS requires a sound model of effective observing behavior.

Artificially Intelligent Tutoring of Shiphandling

COVE ITS detects both process (observation of visual cues) and performance (correct orders at the correct time) as conning officers conduct an assigned evolution in COVE. The ITS also measures students' proficiency, and detects problems that may arise, in ways that we now discuss.

Ship Control Skills

COVE ITS divides shiphandling expertise into a set of fundamental skills, which include controlling the ship's heading, rate of swing, speed ahead or astern, and the lateral speed of the whole ship as well as of its bow and stern separately. These all are controlled by means of a small number of variables: the ship's rudder(s), engine(s) and screw(s), and any other thrusters it might have. These variables interact with each other in complex ways in their effect on the ship's heading and speed(s). Moreover, the variables interact with the ship's speed in their effect on its heading. When a tug is made up with the ship, yet more interactions arise. Because of these interactions, the ITS treats maintaining or regaining a desired trajectory first in terms of the fundamental heading and speed control skills, and only then in terms of combinations of orders for the rudder and engine variables that in the ship's current situation would interact to yield the required heading and speed, or changes thereto. Inexperienced conning officers can make shiphandling errors of either kind: intending to attain a heading or speed that is actually incorrect, or failing to produce a correctly intended heading or speed because of ignorance about engine, rudder, and speed interactions. The ITS must measure both aspects of ship control proficiency (fundamental skills and knowledge of how to produce them) and detect problems of either kind.

Observing Skills

Besides these ship control skills, the ITS must monitor students' observational behavior. Some parts of the environment can be observed from the bridge centerline, others only by standing on one or the other bridge wing. Most displays can be seen only from the bridge itself. In every case, the student's gaze must be directed toward the information that is to be observed. If the target of observation is distant, binoculars may be needed to see it clearly. The ITS determines from the behavioral model described in the preceding subsection what observations a conning officer should make at critical points during an evolution, and also what observations he should make periodically. Since students are instrumented for the COVE simulator to drive a virtual reality visual presentation, the ITS can compare students' observational behavior with the model to measure observing proficiency and detect problems.

When problems with observation or ship control are detected, the ITS follows policies modeled on instructor's reactions and advice from master mariners to give students pointers before problems become unmanageable, but not before they have a chance to see the effects of an error.

Integration of the Prototype Shiphandling Tutor with the COVE Simulator

The prototype ITS is implemented in Java and communicates with the COVE simulator through an application programming interface (API) created by CSC, the producer of COVE. The API allows the ITS to see orders the student gives for rudder, engine(s), and tug(s), what part of the bridge the student is standing on, and what direction he is looking. The API also allows the ITS to monitor the ship's position, heading, and speed, as well as the state of its rudder, engine(s), and tug(s) if one or more is made up. The prototype ITS implements knowledge about interaction effects among rudder, engines, tug, and speed specifically for the DDG-51 class of ships (guided missile destroyers). With shiphandling knowledge and an observational behavior model built into the ITS, the remaining item it requires in order to measure and detect problems with a student's proficiency is information on the evolution that the student is supposed to carry out.

Four lessons or evolutions have been developed so far for the prototype ITS: twisting a ship in place (changing its heading without moving it forward, backward, or to either side); walking a ship (moving the ship sideways without changing its heading or moving it forward or backward); getting underway from a pier (lifting the ship off the pier, taking it out of the slip, and getting underway in the channel); and transiting a series of visual ranges (sailing through a harbor via a sequence of legs, each defined by two aligned markers). At the beginning of an evolution, an instructor (or, in principle, the student) initializes the ITS for the particular lesson to be tutored.

EVALUATION OF THE SHIPHANDLING ITS INTEGRATED WITH THE COVE SIMULATOR IN COURSES AT SWOS

The prototype ITS + COVE system is currently being evaluated at SWOS in the SWO Introduction (SWOI) course for Officer Candidate School graduates and the Advanced Shiphandling and Tactics (ASAT) course for Fleet returning Division Officers. Since March 2011 the study has assessed proficiency gains of students taught by COVE ITS with oversight from an instructor, as compared with proficiency gains of a control group of students who received SWOS's usual one-on-one training from an instructor. All students who are bound for DDG-51 class ships are invited to participate in the study. In the first three trials, 24 SWOI students and a total of 34 students in two ASAT course convenings volunteered (see Table 1). These students were randomly assigned to experimental and control groups.

Table 1: Participant demographics

	SWOI	ASAT
Females/Males	9F/15M	16F/18M
Average age (SD)	25.7 (2.6)	23.5 (1.4)
Average Months at sea (SD)	0	6.8 (3.7)

SWOS selected one lesson for initial evaluation: Getting Underway from a Pier. This evolution subsumes both walking the ship (from the pier to the center of the slip) and twisting it in the channel. All students participating in the study performed this evolution once with instruction from the ITS (experimental group) or a SWOS instructor (control group), and subsequently performed the evolution again without any instruction. Instructors assessed the quality of each student's second evolution according to a SWOS standard for shiphandling. All students in both groups were briefed by an instructor before their own instructed evolution, and likewise debriefed after it—whether they were tutored by the ITS or taught by a SWOS instructor. All also had opportunities to watch other students carry out the evolution.

The ITS recorded data on ship location, speed, heading, engines, and rudder; all orders given to both the helm and tug; location of the student on the bridge or bridge wings; and the direction of the student's visual attention. It also recorded what the ITS said by way of tutoring. On instructor taught runs and on test runs, when the ITS did not speak, the utterances that the ITS would have said were also recorded. Lastly, an audio recorder was used to allow us to analyze how much and what the instructor and other students said during each run.

At the conclusion of their final run all students completed a questionnaire intended to assess their prior knowledge and background.

Data Reduction and Analysis

These three sessions produced several types of data including instructors' evaluations, student behavior (both visual attention and orders), and ship behavior. As the ITS automatically collected both student behavior and ship conditions every second, these performance data were reduced by a Java data extraction module to a discrete number of more meaningful measures. Measures included such things as whether the student gave the orders for lifting off the pier in the correct sequence and other possible errors, how long it took to exit the slip, and crashes or unacceptable proximity to obstacles such as other ships, the quay wall, and buoys, etc. Other measures include tallies of ITS utterances and their types (warnings, directing attention, providing information, etc.).

Due to the quantity of data, transcription and analysis of the audio data was not complete at the time of this writing and will not be reported here.

Results

Instructor scores.
At the conclusion of each run, instructors completed the Conning Officer Shiphandling Assessment (COSA) or Standard Surface Force Shiphandling Assessment, developed by CRESST. This is a weighted set of graded skill attributes with 300 as a perfect score. Instructor scores for the uninstructed test run were not significantly different between the control and experimental groups for either class. For the SWOI class the mean for the control group was 160.83 (SD=67.93) and for the experimental group was 136.33 (SD=66.8). These are not significantly different, $t(18) = 0.75$, *n.s.*. For the ASAT class the mean for the control group was 259.85 (SD=50.54) and for the experimental group was 246.06 (SD=53.09). Likewise, these are not significantly different, $t(30) = 0.75$, *n.s.* (see Figure 1). Naturally the ASAT students, being more advanced, did better than the SWOI students. These scores are somewhat subjective and prone to potential bias because the instructors knew which students were in which group and graded their own students.

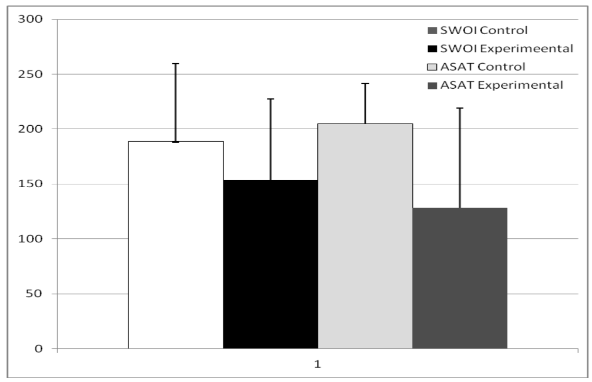

Figure 1. Mean instructors' weighted scores for all groups for the test run.

Behavioral metrics.
In the first, tutored run, 14 of the 23 SWOI students were able to complete the exercise and maneuver the ship safely out of the slip–eight of the control and six of the experimental students. The students of the control group took an average of 337.9 seconds (SD = 25.4) and students of the experimental group had a mean of 359.5 seconds (SD = 95.3). During the second, graded run, seven of the control and nine of the experimental students were able to complete the exercise. These were not necessarily the same students for the first and second runs. Their mean time to exit the slip for the test run was: control group mean = 415.9 seconds (SD = 150.6) and experimental group mean = 397.4 seconds (SD = 63.8) (see Figure 2). None of these times were significantly different.

In the ASAT course, on the instructed run 11 of the 13 control students and 12 out of 15 of the experimental students successfully completed the exercise with a mean time to exit the slip of 497.0 seconds (SD = 76.8) for the control group and 561.3 seconds (SD = 203.9) for the experimental group. For the second, test run all 13 of the control group and 13 out of 15 of the experimental group were successful. Their mean time to exit the slip for the test run was: control group mean =

591.9 seconds (SD = 163.8) and experimental group mean = 482.6 seconds (SD = 110.9). There were no significant differences among the groups on the test run, despite the ASAT control group appearing to be a bit slower than the other groups (see Figure 2).

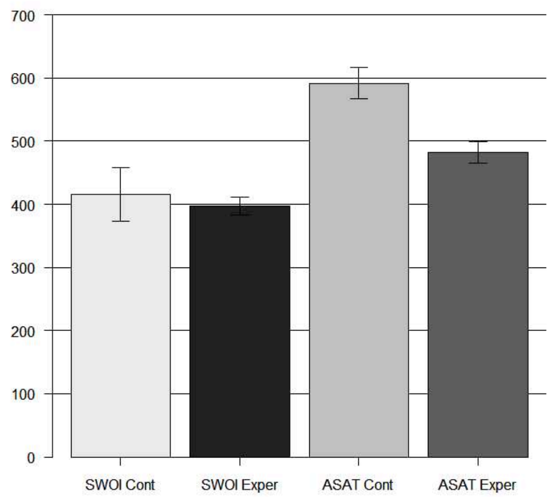

Figure 2. Mean time (in seconds) to exit the slip for all groups for the test run.

Other behavioral metrics included number of orders given, excessive speed, wrong sequence of orders (*not* rudder–engines–tug), bow touching pier, bow off first, crash, minimum distance to other ships and buoys, speed over 4 knots (which triggered an aborted run for experimental students because at that speed the tug cannot function and is in danger), and being substantially off pier heading. Most of these behaviors would result in the ITS providing associated instruction to the student. A composite score was developed that included minimum distance to nearest vessel and buoy, speed outside the recommended range of 1 to 2.5 knots sternway, and getting substantially off pier heading. The highest possible score was 100 for a perfect run; crashing resulted in a score of 0. During the test run the SWOI control group had a mean score of 35.71 (SD = 38.83) and the SWOI experimental group had a mean score of 58.571 (SD = 32.98). Test run scores for the ASAT students were: control group mean = 84.52 (SD = 28.22) and experimental group mean = 79.05 (SD = 26.38). There were no statistically significant differences between the control and experimental groups, although the ASAT students, being more experienced, did have higher scores (see Figure 3).

While it is impossible to prove the null hypothesis, none of the numerous measures we examined provide any evidence that students taught by an instructor achieved greater shiphandling proficiency than students

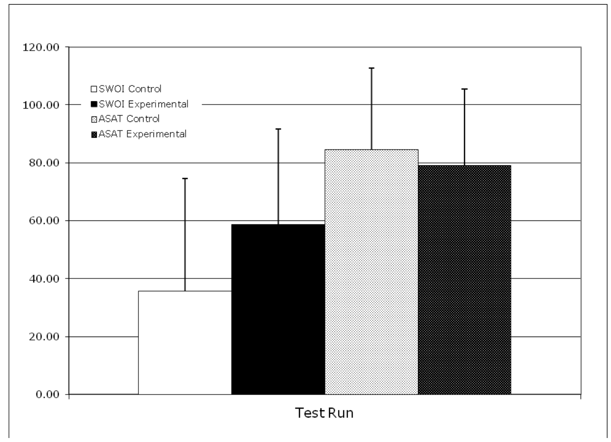

Figure 3. Mean behavioral scores for the test run for all groups.

tutored by the ITS.

CONCLUSIONS AND FUTURE RESEARCH

The results of the evaluation so far strongly indicate that SWOI and ASAT students can be taught shiphandling to the same level of proficiency as they previously have been but utilizing significantly fewer instructor hours by giving much of students' initial instruction with a suitably designed shiphandling ITS. The study has not yet demonstrated how much less of an instructor's time per student will suffice when augmented by an ITS. However, in the near future we expect to test one instructor overseeing two COVE simulators at which the ITS is tutoring students, and to compare these students' performance in their second, uninstructed run with the second runs of control group students whom SWOS instructors taught as usual during the first run. If these experimental and control groups show no significant differences in proficiency during their second runs, then the test would be repeated with one instructor overseeing three COVE simulators at which the ITS is tutoring students. In this way we expect to measure the increase in instructor productivity that a shiphandling ITS can be expected to provide. We are optimistic about the prospects of this planned research, and anticipate that ITSs will indeed prove capable of reducing ongoing costs of simulation-based training very substantially by allowing each instructor to achieve educational goals in less time per student.

A second direction deserving future study is the effectiveness of ITSs in teaching more advanced shiphandling to more experienced conning officers, which SWOS is also responsible for in its Department Head and its Prospective Commanding

Officer/Prospective Executive Officer courses (see Figure 4). For this purpose, the ITS needs enhanced adaptive capabilities to pose usefully challenging problems to students with different backgrounds and experiences, and to adjust the level of difficulty appropriately as students develop over a series of sessions. Another interesting opportunity with these more advanced students is to evaluate the ITS's effectiveness in a 'homework mode', where students choose their own exercises to develop particular skills at which they feel the need for improvement.

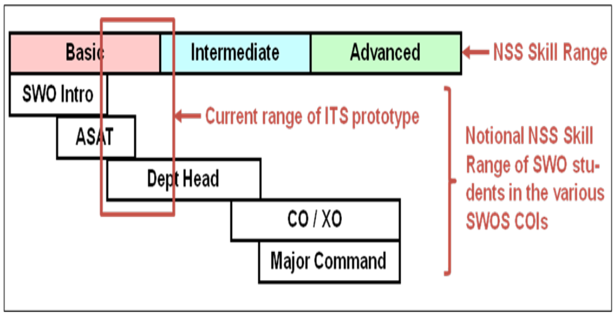

Figure 4. Current and potential range of ITS tutoring

A third worthwhile research direction is to determine the effectiveness of an ITS at teaching beginning shiphandling students who have no instructor available at all. This situation is very challenging for an ITS. If successful, this use of ITSs will be extremely valuable. In many fields, completely naïve students have no way of beginning to learn before going to school to study the subject. Their education would be more efficient, and time in school more productive, if ITSs could provide them the means to achieve a certain base level before arriving at school. In the case of shiphandling, for example, the substantial number of graduating NROTC students would be able to enter the SWOI course ready to learn much of what is now taught in the ASAT course.

ACKNOWLEDGEMENTS

The authors thank the Office of Naval Research for support of this research, the Surface Warfare Officers School for extraordinary cooperation and generous provision of subject matter expertise, and colleagues at Stanford University and the Naval Undersea Warfare Center Division at Newport for numerous valuable contributions to this research.

REFERENCES

Barber, J.A. (2005) *Naval Shiphandler's Guide*. Annapolis: Naval Institute Press.

Crenshaw, R. S. (1975) *Naval Shiphandling*. Annapolis: Naval Institute Press, fourth ed.

Noel, J. V. (ed.) (1989) *Knight's Modern Seamanship.* New York: John Wiley & Sons, Inc, eighteenth ed.

Stavridis, J. & Girrier, R. (2007) *Watch Officer's Guide: A Handbook for All Deck Watch Officers*. Annapolis: Naval Institute Press, fifteenth ed.