The concept of knowledge and how to measure it

Darwin P. Hunt

Human Performance Enhancement, Inc., Las Cruces, New Mexico, USA

Abstract *Knowledge is often defined as a belief that is true and justified. This definition has led to its measurement by methods that rely solely on the correctness of answers. A correct or incorrect answer is interpreted to mean simply that a person knows or does not know something. Such methods of measurement have serious deficiencies that can be alleviated by expanding the definition of knowledge to include the test-taker's certainty. The person's certainty about the answers on a test captures important, but now neglected, dimensions of knowledge. Historical roots of certainty as an essential component of knowledge, and some practical benefits of including it, are discussed. An epistemetric method is described which allows people to indicate "How sure are you?" about the correctness of each of their answers. A computer analysis of the person's answers and self-assessment certainty responses provides multidimensional scores about a person's knowledge that remedy some deficiencies of knowledge assessment and achievement tests now employed.*

Introduction

This paper is about two things:

(1) A definition of personal knowledge, with the aim of providing an expanded concept which, in turn, will allow more productive discussions of assessment, knowledge management, individual and organizational performance and training.

(2) Describing an epistemetric method, i.e. a measurement of knowledge, which is more closely related to the ways motivated people acquire, retain and use knowledge to enumerate, select and execute goal-directed actions at work, in the home and at play.

The paper is focused on tacit knowledge which O'Dell and Grayson (1998, p. 3) indicate is, "... found in the heads of employees, the experience of customers, the memories of past vendors". Knowledge is a concept – like gravity. You cannot see it, but can only observe its effects. Because knowledge is an invisible, intangible asset and cannot be directly observed, many people and organizations do not explicitly recognize the importance of knowledge, in contrast to their more visible financial and capital assets (Sveiby, 1997). Sveiby (1997) suggests that knowledge is invisible because it lacks "a generally accepted definition and a measurement standard".

Sveiby (1997, p. 37) defines knowledge as "a capacity to act"; this makes the important distinction between the behavioral potential, which cannot be directly observed, and the observable performance or behavior. Although there are some who do not distinguish clearly between knowledge and behavior, the

failure to do so prevents the formulation of precise questions about the full process by which individuals and organizations acquire, retain and manage knowledge to perform tasks safely, effectively and at a high quality level.

I have recently become aware of several elements of knowledge management that are of central importance in my own area of primary interest which is the ability of people to assess their own knowledge. This important relation between knowledge and people's ability to self-assess whether they do or do not possess some specific knowledge is not new:

> The Master said, Yu, shall I teach you what knowledge is?
> When you know a thing, to recognize that you know it, and
> when you do not know a thing, to recognize that you do not know it.
> That is knowledge (Confucius, 551-479BC).

The notion that a person's certainty is an essential element of a person's knowledge has been discussed for centuries by philosophers and scientists (e.g. Confucius, c. 500BC, in Streep, 1995; Aristotle, c. 300BC, in Auden, 1970; Polanyi, 1974; Russell, 1948; Ayer, 1958; Quine, 1987). Indeed, the everyday usage of the concept of knowing implies that a higher level of certainty is required to say that one knows something than to say that one believes it to be so. For example one might say, in response to a request for directions to Jones' bakery, "I don't know, but I think that if you walk four blocks north, turn right and go one block, you will find Jones' bakery". As we discuss later, knowing something represents a special class of believing; it is that class of correct beliefs about which one is, and has the right to be, certain. Exactly how certain one must be about a belief to qualify as knowing has been called the boundary problem (Quine, 1987) and will also be addressed later in this paper.

The effects of knowledge on behavior

The knowledge of people greatly affects the safety, effectiveness, comfort and satisfaction with which the goals of an individual or an organization are formulated and attained. Knowledge provides an orderliness to our lives which allows us to conceptualize goals, to anticipate and perceive events, and to respond in accordance with the changing needs, purposes and desires. For example, our perceptions depend both on the data we receive through our senses (eyes, ears, skin, etc.) and the knowledge we possess that allows us to interpret them. Contrary to the popular phrase, "Seeing is believing", it is knowledge, beliefs and needs that structure our perceptions by interpreting the data of our senses. An individual's behavior and performance depend both on the knowledge that has been acquired through learning, practice and experience as well as the sensory receptors and the system of muscles, organs, etc.

The process of acquiring and retaining knowledge (and beliefs) in memory is called learning and is a product of all the experiences of a person from the beginning of his/her life to the moment at hand. Traditionally, learning has

been defined as the relatively permanent modification of the behavioral potential (of an organism) which accompanies practice. The behavioral potential that is modified is the knowledge of a person (or group or any living system, see Miller, 1978). As Ayer (1958, p. 10) reminds us, "To have knowledge is to have the power to give a successful performance, not actually to be giving one". A person can possess considerable knowledge as a result of learning, but such knowledge remains a hidden power until the person uses the knowledge to do something – to perform some task, understand something, make a decision or solve a problem. In spite of its being inaccessible for direct measurement, its power of influence over performance can be overwhelming.

Training programs and schools are so important in our society because they provide formal opportunities to acquire knowledge. As a result of these increased powers acquired as a result of training and education, the person has a potential to perform at a higher level than would otherwise be the case. Of course, the experiences with other people and with the environment influence the kind and amount of knowledge that is acquired and retained by a person. Assessing knowledge prior to testing performance of a complex task has the advantage of detecting and identifying knowledge deficiencies before they are revealed by errors in performance or other near-accident incidents.

To be useful to a person, the knowledge must not only be acquired, but also retained or remembered. It is not enough for instructors and trainers to be concerned only with the acquisition of knowledge by trainees. They must also be concerned with the retention of knowledge so that learners will have the knowledge available to them at later times. If the knowledge is acquired but does not influence behavior and cannot be retrieved from memory, e.g. is forgotten prior to its intended later use, then the earlier learning has failed to attain its instructional purposes. These factors are discussed in more detail in Hunt and Sams (1989).

Defining knowledge
What does it mean when we say that a person knows something? What are the dimensions of knowledge? Our interest here is in knowledge as a characteristic of a person that influences the person's behavioral potential. Since knowledge, itself, cannot be directly observed, it must be inferred from observing performance on a test, e.g. questions designed to determine the beliefs of a person about, say, adding two-digit numbers. Knowledge has been conventionally defined as beliefs that are true and are justified. It is reasonable to think of a "true" belief as one that is in accord with the way in which objects, people, processes and events exist and behave in the real world. However, to avoid the philosophical complexities of the meaning of "true" (Fernández-Armesto, 1997), we will use the term "correct" (instead of true) belief to indicate that explicit and agreed-on criteria, e.g. among scientists, subject-matter experts, text book writers, etc., for determining the correctness of something have been met.

Thus, a belief that is incorrect or false does not qualify to be called knowledge. Furthermore, being correct is not enough. To be called knowledge the belief must not only be correct, but also must be justified. Exactly what evidence is necessary and sufficient to allow a correct belief to be justified has been a topic of discussion (largely by philosophers) for more than 2000 years. Plotkin's (1994, p. 12) elaboration is helpful:

> If I say that I know it is raining, then, for this to be a claim of real and certain knowledge, (1) it must be raining, (2) I must believe it to be raining (merely to say that it is, out of whim, and for it to be raining at the time of the whimsy, would not constitute knowledge that it is raining), and (3) I must be justified in having that true belief. By justify, epistemologists mean that the claim must be justified as reasonable rather than not. For example, I might genuinely believe it to be raining, and it is raining, but my belief may be based on what someone else has told me and that person may be none too reliable. I may even know that my informant is sometimes economical with the truth. "How do you _know_ that it is raining?" I am asked. "Why," I answer, "because so-and-so told me." "Well," say the philosophical judges on this matter, "it is indeed raining, and you clearly believe it to be so doing, but your informant is unreliable and therefore you are not justified in your claim. You don't really know with any certainty that it is raining".

His point that being whimsically correct would not constitute knowledge is relevant to a weakness of common multiple-choice tests in which test takers are given credit for guessed-correct answers.

Certainty and self-assessment
To know something means, at least, that the belief is correct and is justified. For many – probably most – people, to know something requires, in addition, that a person is sure or certain of it. Russell (1948, p. 381) points out, "That all human knowledge is in a greater or less degree doubtful is a doctrine that comes to us from antiquity . . .". He may have had in mind Aristotle's comments (Auden, 1970) which are relevant to the proposal that the certainty of the correctness of one's belief, is an important component of knowledge:

> The proud man … is an extreme in respect of the greatness of his claims, but a mean in respect of the rightness of them; for he claims what is in accordance with his merits, while others go to excess or fall short.

> The man who thinks himself worthy of less than he is really worthy of is unduly humble, whether his deserts be great or moderate, or his deserts be small but his claims yet smaller.

> He who thinks himself worthy of great things, being unworthy of them, is vain. (Aristotle, c. 300BC).

Furthermore Quine (1987, pp. 108-109) points out:

> Knowledge connotes certainty (but) what shall we count as certain? . . . one would hesitate to limit knowledge to the absolutely certain . . . We do better to accept the word "know" . . . as a matter of degree. (Knowledge) applies only to true beliefs, and only to pretty firm ones, but just how firm or certain they have to be is a question . . .

However, methods for determining whether a person recognizes that he/she knows or does not know something have not been incorporated into the

measurement and assessment of people's knowledge, except for an occasional brief outburst of interest, say, in confidence testing. As Quine (1987) said, the question of how certain a person's belief must be for it to qualify as knowledge is a "boundary" problem. Where is the boundary between being certain enough and not being certain enough? For practical purposes of affecting real life behavior, the person must be certain enough so that he/she will use the knowledge to make decisions, solve problems and select/execute actions. This, in turn, suggests that the level of certainty required to qualify as knowledge may be different depending upon the utility or importance of the consequences, i.e. what are the possible benefits if the person's action is correct or the possible costs if the action is incorrect.

The problem of knowledge has been addressed extensively by Ayer (1958). His approach is especially relevant to the topic of measuring knowledge because the component of certainty plays a significant role in his discussions:

> The first requirement is that what is known should be true, but this is not sufficient, not even if we add to it the further condition that one must be completely sure of what one knows. For it is possible to be completely sure of something which is in fact true, but yet not to know it. The circumstances may be such that one is not entitled to be sure (Ayer, 1958, p. 29).

> Suppose that someone were consistently successful in predicting events...like the results of a lottery . . . we might come to say that he knew which number would win. How does our man who knows what the results of the lottery will be differ from one who only makes a series of lucky guesses? . . . so far as the man himself is concerned . . . his procedure and state of mind . . . may be exactly the same as when it is said that he is only guessing. The difference is that to say that he knows is to concede to him the right to be sure, while to say that he is only guessing is to withhold it (Ayer, 1958, p. 31).

This leads Ayer (1958, p. 34) to conclude, ". . . that the necessary and sufficient conditions for knowing that something is the case are first that what one is said to know be true, secondly that one be sure of it, and thirdly that one should have the right to be sure". It seems fair to say that his discussion of the "right to be sure" is quite similar to discussions about the requirement that to qualify as knowledge, a belief must be justified.

Partly as a result of not incorporating the component of sureness into the concept of knowledge, the dominating view in knowledge measurement and assessment is that, if a test taker's answer is correct, then it is considered to be knowledge even if it is of little or no practical use because the person is so unsure of it. If the answer is incorrect it is inferred simply that the person does not know it, which allows misinformation, i.e. being extremely sure that an incorrect answer is correct, to remain unrevealed.

Sureness importantly influences the extent to which people properly utilize their knowledge and beliefs in everyday decisions and behavior. Incorporating the component of certainty into the definition and measurement of knowledge means that a correct justified belief would not qualify as knowledge unless the person is sure of it. To expand the concept of knowledge to include certainty

and to provide a method for its measurement, it is required that there be criteria for determining the correctness of a person's answer and a method to measure how sure the person is of its correctness along with a criterion or boundary above which qualifies as a "sure" response.

A practical way to measure a person's knowledge

How can we measure whether a person knows something? To measure something means to assign a number to a characteristic (knowledge) of an object (a person) or event according to a set of rules. It is the set of rules by which the number is assigned that defines the meaning of the number. The currently used multiple-choice test or any other epistemetric method may be considered as a "set of rules" by which the numbers (scores) or measurements are produced - and thus, knowledge may be operationally defined.

Most tests used today for measuring a person's knowledge on a topic are aimed at composing test items that represent the topic; and are fair and unbiased, i.e. not influenced by the test takers' characteristics other than knowledge, such as gender or ethnicity, which might influence the measurement. To determine whether a person possesses knowledge on, say, simple addition, we can ask questions that are representative of the topic, such as "What is the sum of 12 + 13?"; or we might pose the question as a response selection or multiple-choice task, e.g.

12 +13 = (a) 7
(b) 14
(c) 24
(d) 25

Current testing practice is to observe which alternative a person selects and infer that s/he knows (if a correct answer is selected) or does not know (if the correct answer is not selected) how to add two digit numbers. However, a test taker can select the correct answer without knowing how to add, e.g. in the above example, the chance of being correct by guessing alone is $1/4 = 25$ percent. The reliance exclusively on the correctness of the answer implies that the person who provides a correct but unsure answer or who made a lucky guess possesses knowledge equivalent to a person who is correct and extremely sure of it.

Similarly, in today's multiple-choice tests if an incorrect answer is selected, then it is interpreted simply to mean that the person does not know the answer, i.e. is *uninformed*. This inference is misleading. Specifically, the person may be extremely sure that the incorrect answer which he/she selected is correct and, thus, may be *misinformed*– which is much worse than being uninformed. A sure-but-wrong belief, used confidently as a basis for making decisions and taking actions, may lead to surprising errors in performance – sometimes with tragic results. For example, if a licensing or certification test is being

administered to a professional (say a physician or an aspiring key decision maker), it is important to make the distinction for incorrect answers between a person who:

- is not sure at all as to whether an incorrect answer which he/she gave is correct and thus the incorrect belief is not likely to be employed in practicing the profession; or
- strongly believes that the selected incorrect answer is correct and is therefore likely to use the erroneous belief in making decisions.

Pears (1971, p. 15) remarks about the relation between the level of confidence in a belief and the likelihood of a person taking action based on the belief:

> Think of the person who makes a true statement based on adequate reasons, but does not feel confident that it is true. Obviously, he is much less likely to act on it, and, in the extreme case of lack of confidence, would not act on it.

For instructional guidance as well as qualification testing it is helpful, and may be critical, to detect and identify misinformation and, though difficult, try to remedy it. As Colton (1829 in *The Lacon* in Seldes, 1985) tells us, "malinformation is more hopeless than noninformation; for error is busier than ignorance". Hidden misinformation not only leads to bad decisions and errors in performance, but also is counterproductive as a foundation for more advanced learning.

Some effects of self-assessment on knowledge
As indicated earlier, the close relation between a person's knowledge and certainty has long been discussed. Russell (1948, p. 382) raises the question of whether "there is only one datum, namely a proposition with a degree of credibility attached to it", or "the datum and its degree of credibility (are) two separate data".

For 20 years at New Mexico State University, and recently at Stockholm University, we have been conducting research and addressing the practical implications of people's self-assessments of the correctness of their own responses. The surprising results of some of our first studies (Hunt, 1982; Sams, 1989) indicate that learning is expedited by self-assessment (SA) responding. Our expected result was that the secondary task of assessing the correctness of one's responses would interfere with the primary task of learning.

Some insight into how to interpret the observation that SA responding enhances learning is provided by a signal detection analysis (Green and Swets, 1966) of our finding that learning is affected by the order in which the answer and the SA response are executed (Hunt, 1982). In the signal detection analysis, one can conceptualize the person's SA task as that of deciding whether a "signal of knowing" is present or absent within oneself, e.g. a state of the brain. This analysis assumes that the SA responses of "sure" and "unsure" provide reasonable estimates of the person's decision that the "signal of knowing" is

present or absent. The accuracy of the SA responses is estimated by the calculation of the hit rate (HR) and false alarm rate (FAR) as the conditional probabilities of p (Sure|Correct) and p (Sure|Wrong), respectively.

This analysis reveals that people who first give the answer followed by the SA response are more accurate, i.e. better able to discriminate (d_r) between knowing and not knowing the correct answer, than are those who first give the SA response followed by the answer (d_r = 0.85 vs 0.52 for Answer-SA and SA-Answer, respectively). Furthermore, this greater sensitivity in detecting the "signal of knowing" is due entirely to a higher HR (0.61 vs 0.49) rather than the FAR (0.30 vs 0.30). A higher HR means that, when the answer is correct the probability that the person will say "Sure" is higher for participants in the Answer-SA (p = 0.61) group than for those in the SA-Answer group (p = 0.49). The higher HR is consistent with the interpretation that if a correct response is covertly selected, then its execution helps the learner to confirm its correctness. On the other hand, the finding that the FAR is not affected by the order of response execution suggests that the execution of an incorrect answer has no affect on the accuracy with which a wrong answer is identified by a person as being incorrect.

Related to this is the interesting finding that the correct-and-sure answers are executed approximately one second faster than the wrong-but-sure answers. Arnberg *et al.* (1983) also found some evidence of a resistance to change of sure-but-wrong responses, supporting the "stubborn-error" effect reported earlier by Marx and Marx (1980).

Misinformation and usable knowledge

It is commonly accepted that people behave in accordance with their knowledge and beliefs. The more certain the knowledge or belief then the more likely, more rapid and more reliable is the response. If a person strongly believes something to be correct which is, in fact, incorrect, then the performance of tasks that rely on this erroneous belief or misinformation may likewise be in error – even though the response may be executed with confidence. Thus, from an educational and training point of view it is important to detect and identify misinformation.

This line of thought leads to the concept of "usable" knowledge (Hunt and Furustig, 1989). "Usable", here, means that a person is sufficiently sure of the correctness of the knowledge or belief so that it will be used to make decisions, to solve problems and to select and execute actions. Figure 1 illustrates the relationships among belief, correctness, sureness, usability and knowledge. As indicated earlier, to qualify as usable knowledge, the answer must be correct. If a person strongly believes something to be correct which, in fact, is incorrect, then we will call it a usable belief; and consider the person to be misinformed. If the person strongly believes something to be true, then s/he may use it as the basis for acting – whether it is correct or incorrect. If the belief is correct, then the person may be considered to know it and to be well informed.

Figure 1.
The relationship among the correctness, the sureness or certainty with which the person believes an answer to be correct and usability

Source: Adapted from Hunt and Furustig (1989)

Incorporating certainty as a requirement of knowledge has some important practical benefits, but also raises several problems. Quine (1987) points out one of the problems: How certain must a person be for the belief to qualify as being knowledge? One answer is that the person must be certain enough so that the belief will be used to make decisions, perform tasks and solve problems. If, in addition to that, the belief is correct and justified, then it would qualify as knowledge; otherwise it would not qualify.

Another question is: Since there are different degrees of certainty, what is the most practicable way to indicate these different degrees? For both of the other two qualities (correctness and justification) there are criteria for most practical purposes, e.g. as determined by the subject matter experts, which can establish a belief as being correct and justified or not. For certainty or sureness, it seems more reasonable to define the criterion as a certainty level that is a function of variables such as the benefits (if correct) and costs (if incorrect) rather than a single certainty level.

Self assessment computer analyzed testing

The common multiple-choice test which is widely used in the USA to measure people's knowledge has many advantages which include objectivity, ease and economy of administering and scoring, reliability, and the ability to measure simple and complex knowledge in most content areas at most levels of knowledge. However, the knowledge of a person has more characteristics than is represented by the percentage correct score on a multiple-choice test. Incorporating the concept of knowledge as redefined to include the component of sureness and misinformation into testing produces test scores which are

more representative of the way in which knowledge influences a person's everyday decisions and performance.

The observation that a person recognizes or recalls a correct answer on a test is not sufficient to conclude that the knowledge has been learned, i.e. acquired and will be retained, to a usable level (Figure 1). Similarly, simply by observing an incorrect response one cannot distinguish between a person who is uninformed and one who is misinformed. Self-assessment computer analyzed testing (SACAT) described below provides remedies for both of these inadequacies and has other benefits. It is aimed at providing an epistemetric method that produces measurements more closely related to a person's later performance and is more useful for assessment and instructional purposes than the common multiple-choice test. By detecting and identifying misinformation and providing a measure of the retainability of knowledge that has been acquired, SACAT provides some unique benefits.

The SACAT answer sheet allows the test taker to indicate, "How sure are you", i.e. a level of certainty or doubt, of the correctness of each answer. The additional time usually required to mark "How sure are you?" is minimal (1-5 sec.) since the test takers' assessment of their certainty about the correctness of an answer is often largely completed during the process of selecting or producing the answer.

An "Instructor's Summary" of the test results is provided for test administrators who prefer not to inspect the detailed printouts. In addition to giving summary statements of how well the group of trainees performed on the test, the summary lists for instructional guidance those specific test items about which trainees as a group may be misinformed. On tests in university introductory psychology courses that I taught (with 50-180 students per class), 3-10 percent or more of the test items have a high percentage (> 10 percent) of wrong – but-sure responses indicating misinformation.

The quality of people's performance in real life depends on both the knowledge they possess and the certainty with which they possess it (Hunt and Sams, 1989). With this is mind, a percentage SA (%SA) score, which is an index of the overall accuracy with which a person performs the self-assessments, is reported for each test taker. The %SA scores correspond to the accuracy with which test-takers indicate their certainty when their answer is correct and their doubt when their answer is incorrect. We will call it the Confucius criterion of SA accuracy. The %SA score is calculated by accumulating over all test items, the number of points gained (for correct answers) and lost (for incorrect answers) employing logarithmical gain-loss functions, relative to the maximum number of points possible if perfect sure and not sure assessments of correct and incorrect answers were made.

Being accurate in one's self-assessments is of practical importance and it should be rewarded. To provide such a reward, and to provide an incentive for engaging in self-assessment, the percentage correct (%Correct) score can be increased, say by 3 percent, for the accurate self-assessors. This approach allows

instructors and test administrators who are accustomed to using %Correct scores for grading or qualification purposes, to continue to do so easily.

A different approach is to combine the %SA score with the %Correct score in some weighted fashion dependent on the importance of being correct-and-sure and on the negative consequences of being wrong-but-sure. However, the implications of this approach have not been explored. Jankowicz (1973) has developed a variant of the Strictly Proper Scoring Rules (von Holstein, 1970) for multiple-choice testing, but it has been little used. There are advantages in keeping the %Correct score separate from the %SA score, e.g. to provide feedback to the test taker, until the consequences of combining them are understood.

The SACAT computer analysis of the answer sheets provides a printout of the scores that can be displayed to the test takers for knowledge of results. The test takers who are more accurate in their self-assessment and, thus most deserving of the test bonus, are identified in the computer analysis.

Other features of SACAT
Gender bias
Critics argue that traditional multiple-choice tests are biased against various groups of people, such as women. To be gender biased means, here, that if a male and female know the same amount about the topics of the test, then one of them will obtain a lower score on the test than the other due to a gender characteristic which is not relevant to knowledge. Our own research findings (Hassmén and Hunt, 1994) are:

- female test-takers score lower than males, on the average, when the common multiple-choice test is employed; and
- the difference in the %Correct scores (exclusive of the three-point bonus) between male and female university students is reduced when SACAT is used (Table I).

The benefit of incorporating self-assessment responding may not, however, be limited to women. Koivula *et al.* (2001) recently conducted a study involving Swedish high school and university students to compare performance on the Swedish Scholastic Aptitude Test using either a conventional multiple-choice test or SACAT. For the quantitative subtest they found that those who made

Table I.
Mean number of correct answers for males and females on tests using the usual multiple-choice test and using SACAT

Test used	Gender of participant	n	Number correct	Standard deviation	Sig.
MC	Female	30	23.9	0.54	$p < 0.05$
	Male	30	29.2	0.67	
SACAT	Fem	30	27.8	0.53	ns
	Male	30	29.7	0.55	

self-assessments outscored those who did not, and individuals who rated themselves low on traits traditionally regarded as masculine (measured on a sex role inventory) benefited especially from the self-assessments.

Also, it has been shown that men, especially undergraduate men, were particularly overconfident when incorrect (Lundeberg *et al.*, 1994). In most educational settings this overconfidence has little visible impact on the life and well being of others. However, in many other life situations in which people are dependent upon the judgments and decisions made by overconfident people, the consequences can be quite negative, e.g. operator judgments in process and machine control, automobile passing decisions, etc.

Retainability

The aim of most training is that, as a consequence of the learning that occurs, the trainee will later be able to perform some task or activity. This requires the trainee to acquire the necessary knowledge, skills and attitudes, retain the learned material until a later time when it will be retrieved and used to make decisions, select and execute actions, understand something, etc.

Preliminary findings (Cabigon, 1993) in our laboratory indicate that the retention of responses at the end of learning which are correct are a monotonic increasing function of the level of sureness (Table II). Only 25 percent of the correct responses about which the learner is "Not sure at all" are retained a week later, while approximately 90 percent of the correct responses about which the learner is "Extremely sure" are retained.

This suggests that including self-assessments, along with correctness, as part of the criteria for the assessment of whether a person knows the required material, would improve the correlation between the assessment measures and the quality of later performance. Furthermore, self-assessment responding may help identify the strength of the (stimulus-response) association for both correct and incorrect information.

Motivational effects

On open-ended post-test questionnaires, about 40 percent of the students indicated that they study more to prepare for SACAT, e.g. "to be able to mark that I am sure of my answer", than they do for the usual multiple-choice test.

Sureness on the final learning trail	Percentage retained correctly one week later
"Extremely sure	91
"Very sure"	88
"Somewhat sure"	75
"Very unsure"	75
"Not sure at all"	25

Table II.
The percentage of the correct response on the final learning trail which were retained a week later as a function of the self-assessment level on the final learning trail

Why would students feel motivated to study more when they know that self-assessments will be used during the coming test? One possible explanation is that the process of self-assessment responding acts as a reinforcement for the students, i.e. the feeling (and its anticipation) of being sure during the test is an intrinsic reward thereby motivating the student to study more than otherwise would be the case (e.g., Watson and Tharp, 1993). Since the SACAT method allows extra points to be added to the total score for those students that make the most accurate assessments of their knowledge, this added extrinsic motivation for studying should also work towards increasing the total motivation to study (e.g. Franken, 1994).

Summary of benefits of SACAT
In summary, the benefits of SACAT observed so far are: it provides a more comprehensive measure of a person's knowledge; detects and identifies topics in which people are misinformed; measures the retainability of learned material; may reduce gender and perhaps ethnic bias in the assessment of knowledge; encourages study and enhances learning, identifies and provides practice to individuals who over or under estimate their own knowledge; and helps identify test items which may be misleadingly constructed.

References
Arnberg, P., Fernandez, J. and Hunt, D.P. (1983), "A comparison of learning by doing and learning by observation with self-assessment responding", unpublished.

Auden, W.H. (1970), *The Portable Greek Reader*, The Colonial Press, Clinton, MA.

Ayer, A.J. (1958), *The Problem of Knowledge.*, Macmillan and Company, London.

Cabigon, H.J. (1993), "Effects of self-assessment on retention of training", unpublished Master's thesis, New Mexico State University, Las Cruces, NM.

Fernández-Armesto, F. (1997), *Truth*, Transworld Publications, London.

Franken, R.E. (1994), *Human Motivation*, 3rd ed., Brooks/Cole, Pacific Grove, CA.

Green, D.M. and Swets, J.A. (1966), *Signal Detection Theory and Psychophysics*, John Wiley & Sons, New York, NY.

Hassmén, P. and Hunt, D.P. (1994), "Human self assessment in multiple choice testing", *Journal of Educational Measurement*, Vol. 31, pp. 149-60.

Hunt, D.P. (1982), "Effects of human self-assessment responding on learning", *Journal of Applied Psychology*, Vol. 67, pp. 75-82.

Hunt, D.P. and Furustig, H. (1989), *Being Informed, Being Misinformed and Disinformation: A Human Learning and Decision Making Approach*, Technical Report PM 56:238, 1989-04-21 Institution 56 Manniska Maskin System, Karlstad.

Hunt, D.P. and Sams, M. (1989), "Human self assessment process theory: an eight-factor model of human performance and learning; and everyman's causation", in Ljunggren, G. and Dornic, D. (Eds), *Psychophysics in Action*, Springer-Verlag, Heidelberg.

Jankowicz, A.Z. (1973), "Feedback for learning in business games", *Simulation and Games*, Vol. 4 No. 2, pp. 175-203.

Koivula, N., Hassmén, P. and Hunt, D.P. (2001), "Performance on the Swedish Scholastic Aptitude Test: effects of self-assessment and gender", *Sex Roles*, Vol. 44 No. 11/12, pp. 629-45.

Lundeberg, M.A., Fox, P.W. and Puncochar, J. (1994), "Highly confident but wrong: gender differences and similarities in confidence judgments", *Journal of Educational Psychology*, Vol. 86, pp. 114-21.

Marx, M.H. and Marx, K. (1980), "Confirmation of the stubborn-error effect in human multiple-choice verbal learning", *Bulletin of the Psychonomics Society*, Vol. 16, pp. 477-9.

Miller, J.G. (1978), *Living Systems*, McGraw-Hill, New York, NY.

O'Dell, C. and Grayson, C.J. Jr (1998), *If Only We Knew What We Know*, The Free Press, New York, NY.

Pears, D. (1971), *What Is Knowledge?*, Harper & Row, New York, NY.

Plotkin, H. (1994), *Darwin Machines and the Nature of Knowledge*, Harvard University Press, Cambridge, MA.

Polanyi, M. (1974), *Personal Knowledge*, The University of Chicago Press, Chicago, IL.

Quine, W.V. (1987), *Quiddities: An Intermittently Philosophical Dictionary*, The Belknap Press of Harvard University Press, Cambridge, MA and London.

Russell, B. (1948), *Human Knowledge: Its Scope and Limits*, Simon and Schuster, New York, NY.

Sams, M.R. (1989), "Effects of observational assessments and patterns of success-failure on self-confidence", unpublished dissertation, New Mexico State University, Las Cruces, NM.

Seldes, G. (Ed.) (1985), *The Great Thoughts*, Ballantine Books, New York, NY.

Streep, P. (1995), *Confucius: The Wisdom*, Bowker, Ingram, East Grinstead.

Sveiby, K.E. (1997), *The New Organizational Wealth: Managing and Measuring Knowledge-based Assets*, Berret-Koehler, San Francisco, CA.

von Holstein, S. (1970), *Assessment and Evaluation of Subjective Probability Distributions*, Stockholm School of Economics Research Institute Press, Stockholm.

Watson, D.D. and Tharp, R.G. (1993), *Self-directed Behavior: Self-modification for Personal Adjustment*, 6th ed., Brooks/Cole, Monterey, CA.

Further reading

Hunt, D.P. (1984), *Human Self-assessment and Its Implications for Human Training and Performance*, Summary Report for US Army Research Institute for the Behavioral and Social Sciences.

Hunt, D.P. and Hassmén, P. (1997), *What It Means to Know Something*, Report No. 835 Department of Psychology, Stockholm University, Stockholm.

Kolak, D. and Martin, R. (1996), *Wisdom without Answers: A Brief Introduction to Philosophy*, 3rd ed., Wadsworth, New York, NY.

Reason, J.T. (1990), *Human Error*, Cambridge University Press, Cambridge.