

Methods for Measuring the Influence of Concept Mapping on Student Information Literacy

Carol A. Gordon, Head, Educational Resources Library and Associate Professor, Boston University

Research traditions in education and information retrieval have grown up in parallel worlds, although they share a theoretical foundation that profoundly influences research methodology and best practice in their respective domains. They also share a common problem: the need for a method for analyzing sparse, quantifiable data collected in qualitative studies with small sample sizes. This paper explores the theory of expected information, which uses formulas derived from the Fano measure (1961) and Bayesian statistics (1764), and demonstrates its application in a study on the effects of concept mapping on the searching behavior of tenth-grade biology students.

Common Ground: Research in Education and IR Studies

Behaviorism was defined by Skinner's (1965) theory of operant conditioning, which claimed that behavior could be shaped by reinforcing, or rewarding, desired responses to the environment. Educators became concerned with learning outcomes and devised steps to help learners achieve desired behaviors. Behaviorism also became a theoretical basis for systems approaches in information retrieval (IR) for research that was system rather than user centric. Behaviorists did not make inferences about how learners process information or what goes on internally when learning takes place. In education, behaviorist theory informed practice by providing a rationale for programmed instruction, teaching machines, and computer-assisted instruction. In library instruction it encouraged a tool-specific approach to teaching information skills in isolated lessons taught out of the context of their utility.

Cognitive psychologists investigated learning and created models of how information was received, processed, and assimilated into the learner's knowledge system. Piaget's (1928) theory of cognitive development traced the development of the child. Stages of development included the sensori-motor stage of infancy; the intuitive or preoperational stage of early childhood; the stage of concrete operations in the elementary school years; and the stage of formal operations in adolescence (Inhelder and Piaget 1958). Piaget described schemata—mental structures by which individuals organized their perceptions into categories to classify specific information. These schemata adapt during the learning process through assimilation, by which the learner integrates new information into existing schemata, or by accommodation, whereby existing schemata are modified to create new mental structures. Learners were viewed as actively assimilating and accommodating information in terms of what they already knew (Bartlett 1932; Inhelder and

Piaget 1958). Constructivist theory has grown to provide a rationale for hands-on, active learning; inquiry learning; learning to learn; and performance-based assessment in the classroom. With a paradigm shift from bibliographic instruction to information literacy, this theory supported the process approach to teaching information skills in the academic context of curriculum. The information user is seen as learner through the lenses of information literacy, cognitive, and metacognitive processes.

The behavioral and constructivist schools of thought emerged from educational research that built its knowledge on a philosophical foundation, an ethnographic tradition of observation, and the practical study of human beings. Piaget (1928) and Dewey (1943) observed the child as the object of study as they studied the phenomenon of learning. When the ideas of Thorndike (1903) supplanted this ethnographic approach, psychology as an empirical science became the new foundation for building theory. Thorndike's dictum, "Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality" (Lagemann 2000, 57) prodded educational research to mature into an empirical science. The subsequent contributions of Binet (1916), Binet and Simon (1983), Galton (1883), Pearson (1896), Spearman (1904), and Fisher (1935) established quantitative measures as the dominant, and most credible, kind of data. Lagemann criticizes educational psychologists who embraced mental testing: "But having found a technology that could be applied and tinkered with endlessly, they generally avoided questions concerning the value and necessity of sorting students in the first place" (2000, 94). The Eight-Year Study conducted by the Progressive Education Association moved Tyler (1950) to use learning objectives to guide test construction. Fueled by the large number of college applicants and rooted in education's march to empiricism, the Educational Testing Service was born, creating a culture of standardized testing. With the writings of Lincoln and Guba (1985), qualitative methods began to find a place in educational research. The rebirth of qualitative research and philosophical traditions in educational research, along with the adoption of principles of constructivism from cognitive psychology, provide the common ground where educational research and IR studies can find their roots.

This is significant for their respective research agendas since learning theory adopted as a theoretical framework in a research study has a particularly strong effect on that study's methodology.

The behavioral, system-centric tradition in IR studies grew from a bibliographic paradigm: "Information retrieval has concentrated on what matches the system's representation of texts rather than responding to the users' problems and process of information gathering" (Kuhlthau 1993, 1). These studies, conducted largely by engineers and scientists who were pioneers of IR development, relied heavily on probabilistic theory and algorithmic approaches. User-centric studies, on the other hand, were based on constructivist theory and collected data through ongoing interaction with the user. The system-centric school considered relevance too elusive and subjective and rejected it as a criterion for performance testing. On the other hand, user-centric studies explored relevance in the context of the sense-making approach and experimented with relevance judgments prior to accepting or rejecting relevance as a criterion for performance. The dominant research model of the 1960s and '70s emphasized such input processes and components as document representation and retrieval effectiveness. Learning theories derived from cognitive psychology—which grew from concepts such as knowledge states, conceptual framework, and internal representation—constituted a trend away from system-centric views and shifted the focus from relevance to the information needs of the user. The theory of an

anomalous state of knowledge (ASK) described by Belkin, Oddy, and Brooks (1982) counteracted the best match principle. Dervin and Clark's (1987) sense-making approach and Taylor's (1968) user-value approach emphasized the user's perceptions of the information problem and of the utility and value of the system. Progress in sculpting a theoretical framework based in cognitive psychology culminated in Kuhlthau's (1986) model for the information search process, which used Kelly's (1963) theory of constructs and included the thoughts, actions, and feelings of the information seeker. IR studies used interviews, think-alouds, observation, journaling, concept mapping, and other methods that yielded qualitative data.

Constructivist learning theory and qualitative investigative methods have emerged as powerful research tools that have transformed practice in educational and information skills programs, shortening the distance between the classroom and the library.

The Quantitative-Qualitative Divide

The debate between quantitative and qualitative research is bogus in that the value of their respective methodologies lies not in their relative merits, but in their appropriateness to the research question at hand. There is a tension, however, that arises from practical, rather than theoretical, considerations. At the root of this tension is an overconfidence in what can be quantified and a lack of confidence, or interest, in more cumbersome verbal data.

A practical consequence of qualitative research's lack of credibility is the inappropriate use of quantitative research findings as they are applied to all points of analysis in education—including learner, classroom, and school—to produce results that raise concerns about validity. On a political level, the bias for the quantifiable, which offers a succinct interpretation of data, puts this kind of data in headlines. Although there are appropriate uses for norm-referenced tests, results are often interpreted in terms of individual student achievement. Funding, and even the very existence of marginal schools, may be determined not on how well students are progressing with respect to their own learning history as a baseline, but how well they are progressing compared with everyone else.

The emergence of qualitative assessment measures in education and the teaching of information literacy, namely performance-based (authentic) assessments, are endangered by the absence of a method for handling verbal data quantitatively. These methods use rubrics to determine the attainment of standards as measured by performance ratings that are described at each level of attainment, journals that document process, and portfolios that supply longitudinal evidence of growth. Mounds of qualitative evidence are accumulating devoid of any quantifiable analysis.

Performance-based assessments are intended to be formative—they yield continuous feedback to the learner, providing opportunities for revision in order to improve the performance of both the learner and the teacher. Traditional testing and assessment are intended to be summative, using grades or percentile ranks. While both formative and summative assessment are useful and necessary for the full cycle of the instructional process, the data collected by performance-based instruments cannot easily be pronounced publicly or used politically to garner support and funding. This is a problem for educators, who know that a percentile ranking is a slick and clean way to pigeonhole student performance but offers little substantive information with regard to diagnosis and remediation. Even item analyses of tests, while identifying weaknesses in students performances, are applicable to group performance in terms of a class, grade level, or school but

do not adequately address deficiencies or remedial needs of the individual student. Although performance-based assessment and its attendant constructivist methods are supported by the research, rote learning and behavioral teaching styles that are not information-literacy friendly thrive in a climate of test-driven curricula and the demand of state departments of education for measurable results.

Qualitative research seems to defy definition. Lincoln and Guba wrote, “it is not possible to provide a simple definition. . . . Instead a proper impression . . . can be gleaned only as an overall perspective” (1985, 8). A common criticism of qualitative research is that soft data do not escape the subjectivity of the researcher, raising doubts about reliability. In fact, the analysis of soft data explores the significance, or possible meaning, in the incidence of an event. Qualitative research has an effective tool—the method of constant comparison (Glaser and Strauss 1967)—that is based on the premise that repetitions and patterns in data are meaningful. The significance of this is underappreciated as the bias against lengthy, verbose analyses that defy clear and efficient encapsulation renders vast bodies of meaningful qualitative research inert. When there is a need to process numerical data, researchers often create hybrid studies that marry verbal and numerical data with corresponding qualitative and quantitative methods. If measures were available for sparse quantifiable data, that data could be analyzed and triangulated with qualitative evidence in a way that would preserve the integrity of the ethnographic research. Even sparse data carries some information, and it could be argued that a single occurrence of an event provides important information that the event is possible at all.

Classical statistics requires a large sample, using the philosophy of a bias, probability generator, or urn of black and white balls out there in nature without agreement as to the magic point at which the amount of sampled data crosses from sufficient to insufficient. Not surprisingly, some researchers feel uncomfortable with its methodologies. This may be attributed to a lack of experience with numeric and, in particular, probabilistic concepts. It might also be due to a genuine intellectual concern for matters in relation to notions of sampling and standardized testing, excessive classical emphasis on the refutation of the negative hypothesis, and a feeling that many observations that cannot be quantified in probabilistic terms, or for which there is in classical terms insufficient data, nonetheless do seem to provide evidence that appears intuitively reasonable and cannot in good faith be disregarded. Some of these issues could be better addressed in terms of Bayesian statistics.

The Fano Measure and Bayesian Statistics

The treatment described below derives from the theory of expected information presented by Robson (1974), who combined a derivative of the Fano measure with Bayesian probability theory to treat sparse events. This was probably the first attempt to introduce Bayesian methodology into the biomolecular life sciences, particularly the area now known as bioinformatics. Here there was and is a pressing need for analysis, prediction, and decision making in governing future action, even in cases of sparse data.

Information theory has developed in directions that are of interest as statistical measures or as quantifiable concepts in information science. Generally, information theory is not concerned with confidence levels: you merely have a lot of information or very little. To some interpretations, the amount of information is a kind of degree of confidence. This is analogous to saying that the hypothesis would become acceptable at the 43.0 or 88.5 or 99.75 confidence

level, whatever the data give, rather than say, “reject it because it did not reach the magic number of 95.0.” Such a number is a human artifact. However, a certain threshold of value of information can be used to make a decision.

The Fano measure described the kind of information that most closely relates to perceived meaning or new knowledge is of the general kind:

Information learned about A = What you now note about A – What you knew or expected about A.

The expected probability of A happening anyway can be related to that which is due to pure chance, or the probability of a happening in the general case rather than the circumstance of interest, or simply the probability of what you expected on the basis of well-founded subjective knowledge.

The notion of expected probability being the probability of A occurring by chance is the least controversial, not least because the same principle is used in standard statistics. However, it is quite common that one’s prior knowledge is not simply that something happens by chance. The chance situation is not always the most useful or meaningful basis: there is more hope of finding *Oliver Twist* in a library than in the average room in the average house, for example. It is also possible to introduce a kind of probability that actually relates to utility, such that one will tend to make a decision in a direction that is more profitable or less risky. For example, the decision to look for a book in a library rather than the average room might be reassessed if it cost an enormous amount of money to enter the library or execute a library search.

The information measure can be positive (A has more chance than you would expect), zero (the occurrence of A is just what you would expect, and there is no information), or negative (A has less chance than you would expect). The units depend on the base of logarithms used:

- base 2—binary units, or BITS
- base 10—decimal units, or HARTLEYS
- natural logs (base e)—natural units, or NATS

For formal reasons the use of NATS is recommended for statistics: The natural logarithm is usually the automatic setting on a calculator. Metric units of these—centinats or decinats—are also used. The latter is common with 1 nat = 10.

Fano’s measure is a particular well-defined case of the above information. It is one of mutual information:

$I(A;X) = \log [(\text{number of times A and X are observed together}) / (\text{number of times any A is observed with or without X})]$

This reads as the information relating A to X is equal to the number of times A and X are observed together divided by the number of times A is observed (with or without X).

Probability $P(A)$ is the simple probability or self probability of event A . In the limit of large numbers this is given by $n(A)/N$, the number of times event A is counted divided by N the total sample size. Note that N will be:

$$N = n(A) + n(B) + n(C) + \dots$$

summed over all events A, B, C , etc. For example, $P(\text{blue eyes})$ = probability of occurrence of a person with blue eyes.

Probability $P(A, X)$ is the joint probability, or A and X —the probability that they will be counted together. For example, $P(\text{blue eyes, boy})$ is the probability of counting a blue-eyed boy.

A joint probability could be represented as:

$$P(A, X) = P(A) * P(X)$$

Bayes' refinement of the Fano measure included the introduction of probability $P(X|A)$ is the conditional probability of X on A . For example, $P(\text{blue eyes}|\text{boy})$ is the probability of finding blue eyes, given that the person is a boy. It may be calculated from $P(A)$ and $P(A, X)$ by

$$P(X|A) = P(A, X)/P(A)$$

$P(A|X)$ could be obtained by only counting eye color in the set X , boys. Then

$$P(A|X) = n(A, X)/n(X)$$

The number of blue-eyed boys is divided by the total size of the sample, here the number of boys. Some condition X is, in a sense, always present behind the scenes: It is the set in which you perform your counting. In other words, in statistical sampling you are concerned about the representativeness of X .

Mathematical probability is based on a model that assesses the frequencies of sequences of events. Conditional probabilities provide a refinement of the concept so that particular features of a situation are taken into account when probability is assessed (Parsaye and Chignell 1988). If we rely on a frequency view of probability, however, the more features of a situation we consider, the more unique it is and thus there are fewer previous cases to draw on in estimating the probability. This, in turn, reduces our confidence in the accuracy of the probability assessment.

The Bayesian approach to probability relies on the concept that one should incorporate the prior possibility of an event into the interpretation of a situation (Parsaye and Chignall 1988). Bayes' (Parsaye and Chignell 1988) equation is a special application of conditional probability as described above:

$$P(H|D) = P(D|H) \times P(H)/P(D)$$

$P(D|H)$ is the probability of obtaining data D given (conditional on) hypothesis H , and is the quantity normally measured by statistics. Since it is not a probability of obtaining the hypothesis,

but rather of getting that data given the hypothesis to be true, it is properly called the likelihood. However, it is $P(H|D)$, the probability of the hypothesis being true given the data, that is of interest. To get the latter, we have to know $P(H)$ and $P(D)$. The latter is no problem, it can be chosen so that all the $P(H|D)$ add up to one, the formal requirement for a probability. $P(H)$ however, is the probability of a hypothesis before, or without taking account of, the data D . In classical statistics it is impossible to count, or even give meaning to, such a probability. Hence classical statistics cannot obtain the desirable quantity $P(H|D)$ and rather tortuous reasoning must be used, notably, “I assume H if I can show that the $P(H|D)$ I appear to get is not consistent with H happening by chance.” Bayes got around the problem: probabilities cannot represent biases or trends put there out in nature, they can only represent degrees of belief. Incidentally, since we can hold degrees of belief about anything, it is perfectly good statistics to hold degree of belief about a probability or range of probabilities, for example, $\text{Belief}(P(H)|D)$, which is a kind of probability about a probability, or in classical statistical terms, a probability density.

Bayes’ theorem relates the conditional probabilities of events—it allows us to express the probability $P(A|X)$ in terms of probability of $P(X|A)$, $P(A)$, and $P(X)$. This is important because the probabilities that are available are often $P(X|A)$, $P(A)$, and $P(X)$, but the desired probability is $P(A|X)$.

Information between alternatives is simply one Fano measure subtracted from another, the first being about the probability for some fact or event A , and the second about some fact or event not- A , or the information that A will not occur. This takes in the full weight of available evidence—the information for A and the information against it. Note that the information that data carries about a hypothesis being true is

$$I(H;D) = \log [P(H|D)/P(H)] = \log [P(D|H)/P(D)]$$

The information about the hypothesis being true is what is implied by the above: we should write $H = \text{true}$ every time we write H . This is not the same as the information about the hypothesis being false, where we would write $H = \text{false}$ every time we wrote H above. Fortunately, information against something is negative information for it, so we can subtract the two:

$$\begin{aligned} I(H = \text{true}:\text{false}; D) \\ &= I(H = \text{true}|D) - I(H = \text{false}|D) \\ &= \log [P(H = \text{true}|D)/P(H = \text{false}|D)] - \log [P(H = \text{true})/P(H = \text{false})] \end{aligned}$$

Applying the Theory of Expected Information

The theory of expected information is particularly well-suited for qualitative research that seeks to study an phenomenon in depth by reaching for underlying explanations. The theory also fits the needs of local studies in education and in all areas of study where the field in which human behavior is observed is the laboratory where data collection occurs. Implicit in the verbal data generated by such investigations are occurrences that can be quantified, albeit with sample numbers that do not satisfy the prerequisites of quantitative research. Such a study is described in this paper in order to illustrate the application of the theory of expected information to such

quantities, as well as to model how the information measures, which are a kind of statistic, can guide the analysis of verbal data.

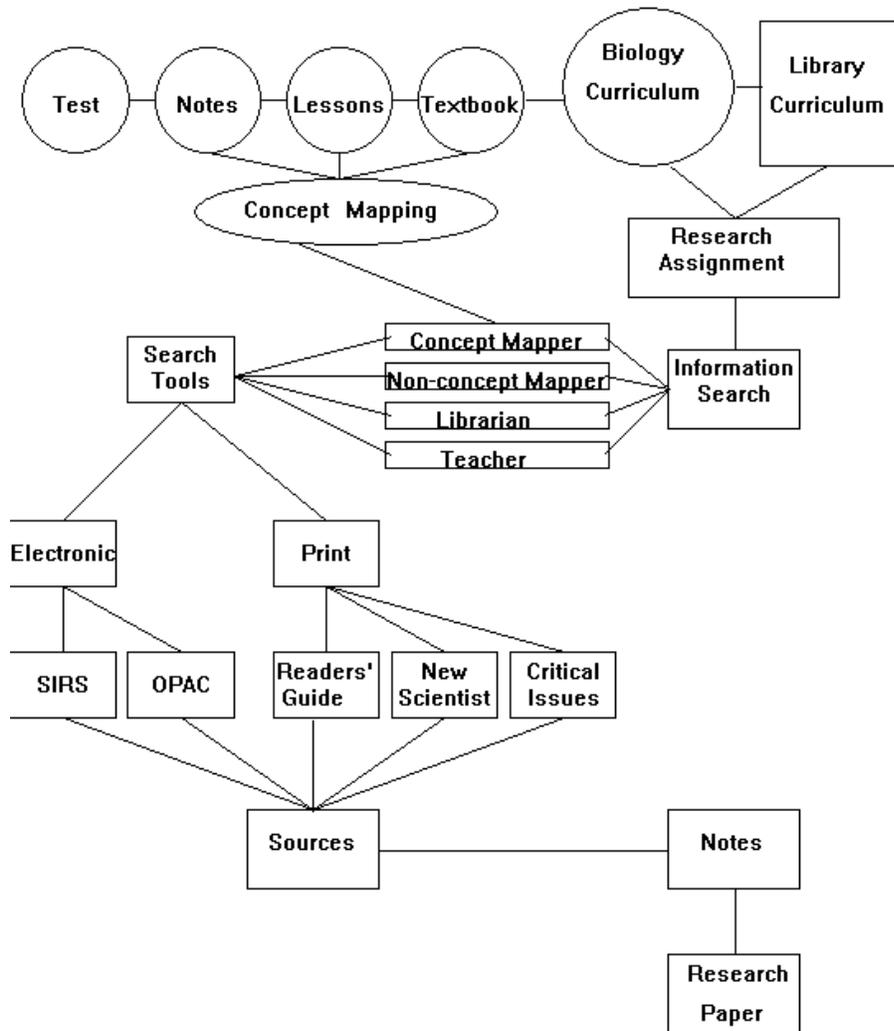
The Research Design

A qualitative study addressed the effect of concept mapping on the searching behavior of tenth-grade students engaged in research projects based on their instruction in a classroom-based genetics unit (Gordon 1995). The setting was an automated library of a private American school in Europe. Ten students were chosen by purposive sampling. Selection criteria, monitored by user profiles, included student age, computer experience, native language, grades, and test scores. How did tenth-grade biology students who learned and used concept mapping in the classroom for seven months search for information in the context of a library research assignment? Research questions included:

1. How did students mappers and nonmappers search for information in the context of the same library research assignment?
2. How did concept formation influence search strategies and relate to developing search strategies?
3. How did expert searches of the librarian and teacher compare with each other and with the searching of mappers and nonmappers?
4. How did student searching relate to stages of the Information Search Process? (Kuhlthau, 1993).

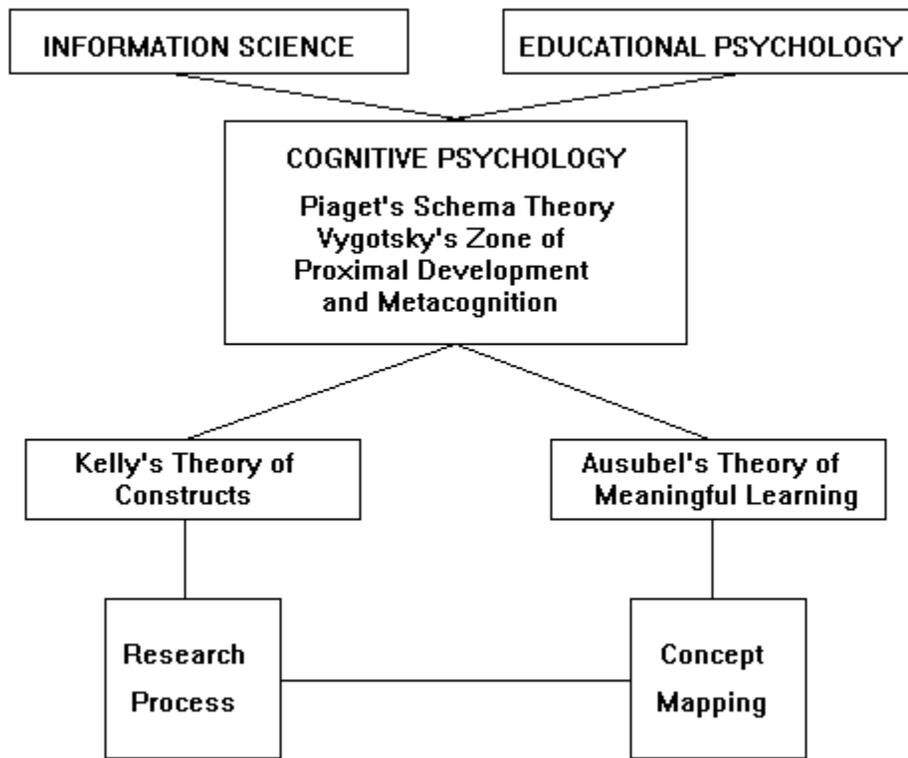
The conceptual framework for the study, shown in figure 1, illustrates the key constructs of the study. One group used concept mapping over a period of seven months, while the other, taught by the same teacher, received the same classroom instruction without mapping. Data on the searching behavior of students, the biology teacher, and the librarian were collected during audiotaped, think-aloud search sessions followed by structured interviews. Participants wrote transaction logs from memory immediately following each search. Stimulated recall was used as an interview method—key informants were provided with written transcriptions of their think-aloud sessions. Participants maintained journals from the beginning of the genetics unit to the end of the research project. Debriefing took place immediately following each session of data collection. The constant comparative method of analysis was applied to these data. Calculations based on Bayesian statistics and the Fano measure from information theory were triangulated with qualitative analysis of data. Information searching, as defined by Kuhlthau's ISP model, was examined to include stages from prefocus formulation to writing the research paper.

Figure 1. Conceptual Model of the Study



The theoretical framework drew from the research traditions in education and information science studies based on cognitive psychology. Figure 2 illustrates the components of this framework.

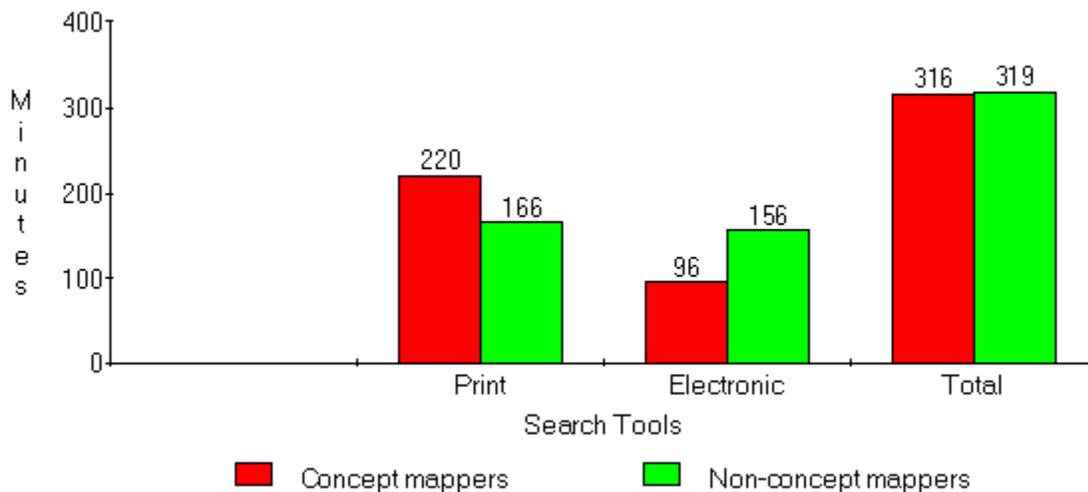
Figure 2. Theoretical Framework of the Study



Analysis of the Data

The raw data for time spent searching on electronic and print search tools is shown in figure 3. The total time spent searching by the two groups differed by only three minutes, with the mappers searching a total of 316 minutes and the nonmappers searching for a total of 319 minutes. There was some difference between the total time spent by the two groups researching print indexes: mappers searched for 220 minutes, while nonmappers searched 166 minutes.

Figure 3. Total Time Spent Searching on Print and Electronic Search



When the time allotted to print indexes was compared with computerized search tools, it was observed that mappers spent 96 minutes of their total time searching electronically, while nonmappers spent 156 minutes on computers. As noted, there was virtually no difference in the total searching time of both groups. Another way of highlighting the differences in apportionment of time is shown in figure 4 for mappers and figure 5 for nonmappers. Mappers spent 6.6 percent of their search time on OPAC, while nonmappers spent twice as much time on OPAC. The total time on computer search tools, such as OPAC and SIRS, was 30.3 percent of total time spent searching by mappers and 48.0 percent for nonmappers. The allotment of time to print indexes was 69.7 percent for mappers and 52 percent for nonmappers.

Figure 4. Search Time Allotment—Concept Mappers

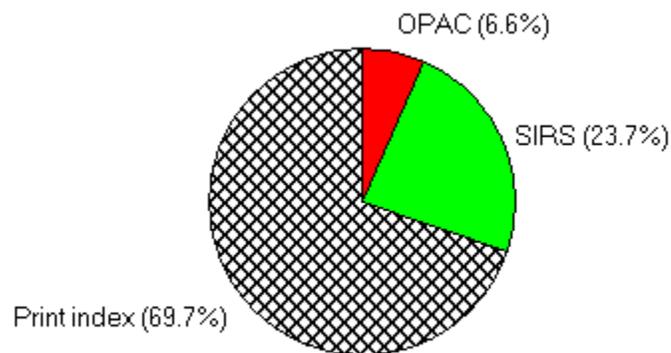
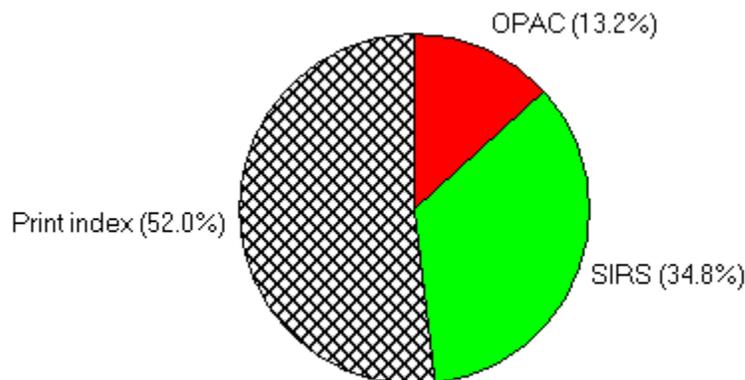


Figure 5. Search Time Allotment—Non-concept Mappers



A useful way of encapsulating the essence of diagrams such as those in figures 3, 4, and 5, and avoiding the subjectivity inherent in visual judgment of graphs, is to process the quantities represented to information measures. Even more important, the approach allows account to be

taken of the fact that a particular type of education or training might have a beneficial effect in favor of some activity A, even if the student still spends less time in that activity A than in other activities. To allow this analysis, the amount of time spent in each kind of pursuit (use of print, use of electronic) is interpreted as proportional to a probability; for example, the probability that the student will be found in the specified pursuit, such as use of print, at any time. The ratio of time in one activity as opposed to another is, in effect, compared with the ratio that we expect if we pool (and thus choose not to distinguish) concept mapping from non-concept mapping students. Logarithms are taken consistent with the theory of information posed by Fano, and in this case natural (base e) logarithms resulting in natural units, or nats, of information. The formula used is as follows:

$$I(\text{use}=\text{print} : \text{electronic} ; \text{mappers}) = \log [P(\text{mappers, print}) / P(\text{mappers, electronic})] - \log [P(\text{print}) / P(\text{electronic})]$$

This could be read as “the amount of information provided by concept map training that the student will use print at any time as opposed to electronic means.” The semicolon means “information provided by” and the colon means “as opposed to.” Substituting the times, we obtain

$$\log [220/96] - \log [386/252] = +0.40 \text{ nats}$$

as the measure of that information. Because we are estimating probabilities from data in regard to time spent, it is customary to express results as being conditional on that assumption about data. To do this, a vertical bar is used to express the notion of conditional on. That is, we may write

$$I(\text{use}=\text{print} : \text{electronic} ; \text{mappers} | \text{time spent}) = +0.40 \text{ nats}$$

This means the information, being positive, is in favor of the student using print means at any time. Repeating this, but now replacing the values for mappers by those for nonmappers, we obtain the information measure

$$I(\text{use}=\text{print} : \text{electronic} ; \text{nonmappers} | \text{time spent})$$

This could be read as “the amount of information provided by non-concept map training that the student will use print at any time as opposed to electronic means.” This yields

$$\log [166/156] - \log [386/252] = -0.36 \text{ nats}$$

That is, we may write

$$I(\text{use}=\text{print} : \text{electronic} ; \text{nonmappers} | \text{time spent}) = -0.36 \text{ nats}$$

This measure, being negative, indicates that the information is against the nonmappers using print at any time. As it happens, the effect is quite strong, such that this can be taken as equivalent to the fact that mappers will be more likely to use print. However the subtlety of the measure is that it could be that the concept map training was shown to have a positive effect even if it was still less probable that print would be used. For example, your vote is a step in the right

direction for your party even if your party doesn't get elected. Note also that basic information theory does not make statements about significance, only about the amount of information available, and represents a weight of evidence, loosely analogous to stating the level of confidence limit at which a given hypothesis would become acceptable in classical statistics. The above measures correspond to the kind of value one would obtain if there was expected to be no bias and the results showed mappers to spend 50 percent more time in print than in electronic, which in this example is consistent with the impression easily gained from the diagram. In contrast to classical statistics, the information theory approach likes to consider both Type I and Type II errors as of relevance, and this is implied in the above approach.

Although not used by Fano in that it loses comparative detail in relation to what is expected, subtracting the second (non-concept mapper) measure from the first (concept mapper) measure is some measure of the total information available in distinguishing mappers and nonmappers, and is

$$I(\text{use}=\text{print} : \text{electronic} ; \text{mappers} : \text{nonmappers} | \text{time spent}) = +0.767 \text{ nats}$$

The difference between measures for mappers and nonmappers is +0.767, or about three-fourths the value of 1 nat (1 = natural log 2.718) or 2.04. This means that if there was the expectation that no bias existed between the use of print as opposed to electronic by mappers and nonmappers regarding time spent, the difference between the time spent by mappers and nonmappers in print is more than 2 nats. A measure of 1 nat or more is generally regarded as particularly meaningful.

An alternative measure is

$$I(\text{user}=\text{concept mapper} : \text{non-concept mapper} ; \text{print} | \text{time spent}) = +0.29 \text{ nats}$$

and

$$I(\text{user}=\text{concept-mapper} : \text{non-concept mapper} ; \text{print} | \text{time spent}) = -0.48 \text{ nat}$$

This measure has some conceptual advantages in that it is predictive, reading as the “information that the time spent in the technique used tells about whether the user is a concept mapper or non-concept mapper.” In the first case, the log of mappers' time in print (220 minutes) divided by nonmappers' time in print (166 minutes) minus the log of total time spent searching by mappers (316 minutes) minus total time spent searching by nonmappers (319 minutes) yields an information measure of +0.29, which measures the amount of information in favor of the user being a concept mapper. Similarly, the second measure of -0.48, obtained by subtracting +0.29 from +0.77, or the difference between the two information measures for mappers and nonmappers of time spent in print as opposed to electronic (+0.40 and -0.36) yields a measure of the amount of information in favor of the user being a non-concept mapper, -0.48. The measure of +0.29 is a useful, single measure, indicating by the positive sign of the measure of mappers that the mappers are doing better. Not least, the nonmappers are the whole complementary set to the mappers (they are mutually exclusive), which covers all possible outcomes and so facilitates interpretation. However it leads to generally similar conclusions but seems less natural with regard to sequence of events in time, such as considering the information provided for the posteriori process of searching by the prior process of the choice of the method of teaching

(concept map or non-concept map). Further, for related reasons we are interested in predicting the search performance of a searcher given whether he is a concept mapper or not, which is another way of saying whether we consider the concept map training useful.

The method used here differs from the Fano measure in that it involves the difference between two such measures, in which the researcher considers the information for a hypothesis or an event then subtracts the information against the hypothesis or event, this latter information being regarded as relevant, but contradicting, information. This resembles other measures, such as the K-statistic. It was then regarded that many such measures could be held for a given set of sparse data, including cases like zero or one observations. A Bayesian expectation of the method was then evaluated over all the different measures that might exist in the mind of the researcher using Bayesian theory. As it happens, this approach and the mathematical integration implied in it led to a quite simple approach with a flavor of its own that allowed a pencil and paper evaluation of the information content of sparse events.

Triangulation with Qualitative Data

Qualitative data provided evidence for formulating hypotheses that could explain the effects indicated by application of the measures of expected information. When asked whether they preferred print or computer searching, mappers did not share a consensus as a group. It was not yet obvious why mappers allocated more time to print.

Nonmappers unanimously expressed a preference for electronic rather than print because computers seemed to make searching easier. Their reasons for the preference indicated that either the computer helped them to cut through the large quantities of information or it provided them with a lot of information, which made them feel confident about searching even if that information was irrelevant.

Data in figure 6 show responses of the ten searchers when asked to identify their best and worst searches. While a stated preference for a search tool is not synonymous with judging that a search was successful, one would expect to find some agreement between the two. Four mappers chose print indexes as their best searches although they had stated they preferred electronic searching. Four nonmappers chose computer indexes as their best searches and print indexes as their worst searches. These data agree with the nonmappers' stated preferences for electronic searching. There was more agreement between the nonmappers' stated preferences for electronic and their choices for their best searches than for the mappers, who said they liked electronic searching but chose print indexes as their best searches. The choices of the two groups for best and worst searches are almost mirror images of each other: print searches were chosen for the mappers' best searches and for the nonmappers' worst four out of five times. Similarly, non-concept mappers chose computer search tools as instruments of their best searches, while concept mappers chose them as instruments of their worst in almost the same ratio.

Figure 6. Searchers' Best and Worst Searches

CONCEPT MAPPERS' BEST	NS (P)	NS (P)	NS (P)	NS (P)	SIRS (E)
NON-CONCEPT MAPPERS' WORST	NS (P)	NS (P)	NS (P)	RG (P)	SIRS (E)
NON-CONCEPT MAPPERS' BEST	SIRS (E)	SIRS (E)	OPAC (E)	OPAC (E)	RG (P)
CONCEPT MAPPERS' WORST	SIRS (E)	SIRS (E)	OPAC (E)	RG (P)	CI (P)

P = Print Indexes

E = Electronic Indexes

NS = New Scientist Index

OPAC = Online Public Access Catalog

RG = Readers' Guide Index

SIRS = Full-Text Database

CI = Critical Issues Index

Scrutiny of the search tools offered some clues for preferences. From the four mappers who preferred print indexes, three stated the *New Scientist* as their preference, which was the most difficult search tool to use because of its complicated coding system. It also indexed the most difficult reading material. The fourth concept mapper who preferred the print index stated a preference for *Critical Issues*, which indexes articles from the *New York Times*. This index was also difficult to use because it was coded to microfiche retrieved from a binder separate from the index. Searchers also had to learn how to use the microfiche reader printer. Mappers did not perceive any of these difficulties as obstacles. Some even saw the print indexes as an aid in focusing: "The *Readers' Guide* has helped me focus by showing me what articles are available on different subjects and to break down the subject." On the other hand, nonmappers chose the same types of indexes as their worst search tools, with the exception of one, who noted in his journal that the *Readers' Guide*:

. . . has been my longest search so far. What I achieved through the search . . . is that out of the huge number of sources the search provided, it helped me to realize that a lot of the information that I had been considering was irrelevant given the time and length limits placed on this paper. I think this was the first search that has made me realize that I must maintain my focus if I am ever going to get this done!

Mappers tendency to spend more time in print indexes was directly related to:

1. the high quality of scientific, technical information retrieved through print indexes;
2. the large number of sources the indexes yielded; and
3. the rich supply of search terms the indexes offered by cross-references and titles.

A closer look at what the searchers were doing with their time revealed more information about the reasons for their preferences and time allotments. The number of sessions that the two groups searched differed, as shown in table 1. Mappers searched a total of 24 sessions, 7 of which were electronic searches, which represents 29 percent of their total number of search sessions, compared with the nonmappers, who had 9 electronic search sessions, or 53 percent of their total number of search sessions. For print sessions, the reverse was true, with mappers spending 17 sessions, or 71 percent of their total number of sessions, in print; nonmappers spent 8 sessions in print, or 47 percent of the total number of sessions. These comparisons support earlier findings that mappers showed a preference for print index searching, allocating more sessions and longer periods of time to them.

Table 1: Number and Length of Search Sessions

	Total time	Mean length of session	No. of sessions	Mean time per searcher
Electronic				
Concept mappers	96	19.2	7	2.7
Non-concept mappers	156	30.6	9	3.7
Print				
Concept mappers	220	12.9	17	1.8
Non-concept mappers	166	20.8	8	2.6
Total				
Concept mappers	316	13.2	24	12.9
Non-concept mappers	319	18.8	17	20.8

When the number and length of search sessions is examined (see table 1), it can be seen that mappers spent more time and more sessions in print, but their sessions were shorter than nonmappers, indicating that they were either searching faster and possibly doing more, or searching shorter periods of time during which they were doing less.

Why did mappers search about one-third of their time on electronic tools while nonmappers searched almost half their time on computers? Why did mappers prefer print indexes while nonmappers spent considerably more time on electronic searches? Did electronic methods compensate for nonmappers' lack of conceptual maps? Do electronic methods benefit mappers or nonmappers?

Critical to answering these questions was the data that would indicate whether mappers were searching more efficiently in electronic tools and more thoroughly in print indexes as compared to their nonmapper counterparts. Examination of the remaining search characteristics using the theory of expected information was intended to explore how searchers used their time and the rates at which they performed search functions in order to make judgments about efficiency and thoroughness of their searches.

These questions led to applying the expected information measure to ten search characteristics.

Findings

Table 2 summarizes the information measures for ten search characteristics. In each instance, supporting qualitative data supplied explanations and insights about these measures.

Table 2: Summary of Measures of Search Characteristics

Search Characteristic	Concept Mappers	Non-concept Mappers
Print vs. electronic	+0.40	-0.36
SIRS vs. OPAC	+0.96	-1.65
Subject vs. key word	+0.74	-0.41
Total search word repertoire	+0.91	-0.74
Unique search word repertoire	+0.61	-0.41
Opening moves	+0.72	-0.52
Reformulations	+0.78	-0.57
Search operations	+0.79	-0.64
Breadth searching	+0.60	-0.47
Depth searching	+0.22	-0.21

The differences between the two measures for each search characteristic is the measure of the probability that mappers searched more thoroughly and more efficiently. In all cases that probability measured at least half a nat (one nat being equal to 2.718), indicating that the chances were approaching twice as likely that searchers engaged in these functions in print indexes were mappers. The plus or minus sign is an indication; conclusive statements cannot be made on the strength of the effect. Since the study defined successful searching in terms of these indicators, findings strongly indicated that the amount of information available was in favor of mappers doing more in their searching; for example, being more thorough searchers who showed a preference for print search tools.

There was a greater probability that mappers will use print rather than electronic means, that they will search in SIRS rather than the OPAC, and that in electronic searching they will conduct subject heading rather than keyword searches. In print, as opposed to electronic searching, measures showed mappers applied a larger number of search terms; employed opening moves, reformulations, search operations, and relevancy judgments more often; and executed more depth than breadth searching. In all cases probability measured at least half a nat, indicating chances were approaching twice as likely that searchers exhibiting these characteristics in print indexes will be mappers. Larger differences between the groups emerged in electronic searching, where mappers spent less time. Quantitative data verified mappers were more thorough and efficient, reformulating by shifting synonyms and moving from general to specific search terms, and terminating searches to read rather than when they depleted their search terms. Stronger focus formulation emerged as the most important determinant of searching behavior. Further research is recommended to replicate the study with a larger sample, using information theory as an alternative to classical statistics in hybrid qualitative-quantitative studies.

A summary of information measures on the number of times per minute opening moves, reformulations, and search operations were performed is reported in table 3.

Table 3: Summary of Measures of Rates of Search Operations

Search Operations	Mappers	Nonmappers
Opening moves	-0.25	+0.30
Reformulations	-0.26	+0.34
Search operations	-0.33	+0.38

Data indicated a trend: mappers will perform these operations about the same number of times per minute and make more relevancy judgments while searching in print as opposed to electronic search tools, which is appropriate given the nature of manual versus electronic searching. With attention to sign, negative values were computed for print as opposed to electronic for reasons of consistency and comparison. The resulting measures indicated that the signs of the computations were consistently in favor of the probability that mappers would use these operations in electronic searching as opposed to print, the sign being inverted for the calculation of electronic as opposed to print. These characteristics were calculated per minute, indicating rate: mappers were performing the functions faster if they were able to do more per minute. Since the study defined the search in terms of these indicators, the findings strongly indicated that from these measures, the amount of information available was in favor of the mappers doing less per minute in their print searching; for example, being more thorough given the nature of print searching.

The probability of assessing whether these results could arise by chance was actually quite difficult for a calculation of this nature. The nature of the indication of an information measure reflects the probability that a concept mapper will get one kind of result as opposed to another. However, the proper procedure would be to progressively increase the number of individuals in the sample until the results converge and become approximately independent of the sample size. This could be accomplished by replicating the study and using cumulative data in order to keep the sample small.

The relationship between the quantification of data and the texture and depth of understanding provided by qualitative data was synergistic in the analytical process that sought to explain as well as describe what was going on in the search process. Emerging patterns of searching behavior in print and electronic environments were outlined by numerical summaries; texture and color of related qualitative description served to provide a measure of understanding. For example, while characteristics of searching behavior—such as search word repertoire, opening moves, and reformulations—could be quantified to point out differences in print and electronic environments, qualitative data provided explanation for relevancy judgments and connections. The interplay of numerical and verbal descriptions served to push the analysis toward understanding the searching phenomenon in depth; for example, the interdependence of focus formulation and information overload. Qualitative data, through triangulation of students' testimony about their own thought processes with observations of their performance, shed light on the metacognitive aspects of searching. Quantification of the data provided direction and structure, as illustrated by the examination of key word and subject searching and search word repertoire, which highlighted characteristics of the concept-driven search. In every instance qualitative evidence supported the findings described through the information measures: mappers were more thorough and efficient in their searching, more inclined to concept-driven searching as evidenced by their ability to focus and make connections, and more inclined to make metacognitive judgments that led to successful searching.

The searching behavior of mappers was more thorough and more efficient if they:

- reformulated by shifting synonyms and moving from general to specific search terms rather than by changing concepts, so their reformulations were within the focus of the search;
- avoided information overload;
- made more connections in a balanced and eclectic pattern from print and electronic search tools; and
- terminated searches because they wanted to read and not because they exhausted their repertoire of search terms.

Analysis of the data has shown that the most dramatic differences between mappers and nonmappers emerged in electronic searching where the mappers:

- spent less time searching;
- searched for fewer and shorter sessions;
- preferred subject heading to key word searching;
- spent less time in OPAC than in the electronic index;
- had fewer search words in the repertoire of nonrepeated words and in total number of search words;
- generated fewer opening moves and generated them at a faster rate;
- generated fewer search strings and generated them at a faster rate;
- generated fewer reformulations and generated them at a faster rate;
- generated fewer search operations and generated them at a faster rate;
- generated fewer relevancy judgments; and
- performed a larger percentage of depth rather than breadth searches.

Based on these observations the researcher concluded that mappers were more sensitive to the electronic environment. They were more efficient in the way they used their time to perform more search operations per minute and more thorough in consistently applying a more concise repertoire of search terms and in engaging in more depth searching. Without the benefit of quantification and the resulting predictive nature of the information measures, the study's findings would have been based on raw numerical data without benefit of degree of certainty about conclusions reached and without the strength of a predictive information measure.

References

- Bartlett, F. C. 1932. *Remembering*. Cambridge, England: Cambridge Univ. Pr.
- Bayes, T. 1764. An essay towards solving a problem in the doctrine of chances. In E. S. Pearson and M. G. Kendall, *Studies in the history of statistics and probability*. London, 1970.
- Belkin, N. J. , R. N. Oddy, and H. M. Brooks. 1982. ASK for information retrieval. *The Journal of Documentation* 38, no. 3: 61–71.
- Binet, A. 1916. New methods for the diagnosis of the intellectual level of subnormals. In E. S. Kite, trans., *The development of intelligence in children*. Vineland, N.J.: Publications of the Training School at Vineland. (Originally published in 1905 in *L'Année Psychologique* 12: 191–244.
- Binet, A. and T. Simon. 1983, reprint. *The development of intelligence in children*. Salem, N.H.: Ayer.
- Dervin, Brenda and Kathleen Clark. 1987. ASQ: Asking significant questions: Alternative tools for information need and accountability assessments by libraries. ERIC 286 519.
- Dewey, J. 1943. *The School and society*. Chicago: Univ. of Chicago Pr.
- Fano, R. 1961. *Transmission of information*. New York: John Wiley and Sons.
- Fisher, R. A. 1935. *The design of experiments*. Edinburgh: Oliver and Boyd.
- Galton, F. 1883. *Inquiries into human faculty and its development*. London: MacMillan.
- Glaser, B. G. , and A. Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.
- Gordon, C. 1995. Concept mapping as a pre-search activity in the research process. Unpublished doctoral dissertation, Boston University, Boston.
- Inhelder, B., and J. Piaget. 1958. *The growth of logical thinking: From childhood to adolescence*. New York: Basic Books.
- Kelly, G. A. 1963. *A theory of personality: The psychology of personal constructs*. New York: Norton.

Kuhlthau, C. C. 1993. *Seeking meaning: A process approach to library and information services*. Norwood, N. J.: Ablex.

Kuhlthau, C. C. 1986. *Facilitating information seeking through cognitive modeling of the search process*. New York: Rutgers Univ. School of Communication, Information, and Library Studies. ERIC, ED 328 268.

Lagemann, E. C. 2000. *An elusive science: The troubling history of education research*. Chicago: Univ. of Chicago Pr.

Lincoln, Y. S., and E. G. Guba. 1985. *Naturalistic inquiry*. Beverly Hills, Calif. : Sage.

Parsaye, K., and M. Chignell. 1988. *Expert systems for expert*. New York: John Wiley and Sons.

Pearson, K. 1896. Mathematical contributions to the theory of evolution, III: Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London* 187: 253–313

Piaget, J. 1928. *Judgment and reasoning in the child*. New York: Harcourt.

Robson B. 1974. Analysis of code relating sequences to conformation in globular proteins: Theory and application of expected information. *Biochem Journal* 141, no. 3: 853–67.

Skinner, B. F. 1948. *Walden Two*. New York: Macmillan.

Spearman, C. 1904. General intelligence: Objectively determined and measured. *American Journal of Psychology* 15: 201–93

Taylor, Robert S. 1968. Question-negotiation and information seeking in libraries. ERIC, ED 015 764.

Thorndike, E. L. 1903. *Educational psychology*. New York: Lemcke and Buechner.

Tyler, R. 1950. *Basic principles of curriculum and instruction*. Chicago: Univ. of Chicago Pr.

School Library Media Research (ISSN: 1523-4320) is the successor to *School Library Media Quarterly Online* and the predecessor to *School Library Research*, an official journal of the American Association of School Librarians. The purpose of *School Library Media Research* is to promote and publish high quality original research concerning the management, implementation, and evaluation of school library programs. The journal also emphasizes research on instructional theory, teaching methods, and critical issues relevant to the school library profession. Visit the [website](#) for more information.



AMERICAN ASSOCIATION
OF SCHOOL LIBRARIANS
a division of the American Library Association

The mission of the American Association of School Librarians is to advocate excellence, facilitate change, and develop leaders in the school library field. Visit the [AASL website](#) for more information.