The Clinical Interview: Validity, Reliability and Diagnosis

DAVID SWANSON, ROBERT SCHWARTZ, HERBERT GINSBURG, NANCY KOSSAN

In a previous paper, Ginsburg, [1981] presented an analysis of the clinical interview procedure as employed in psychological research on mathematical thinking He argued that the clinical interview has legitimate uses for two different research purposes — the *discovery* of cognitive processes and the *identification* or *specification* of cognitive processes When the discovery purpose is stressed, the clinical interview procedure begins with a fairly open-ended mathematical task (e.g an arithmetic word problem) The interviewer questions the subject in a manner contingent upon the nature of the problem solving in which the subject engages The subject is typically asked to reflect aloud, describing the problem solving steps taken and the reasons for them When specification (identification) is the reason for the research effort, the details of the cognitive activities are of interest: the structure, processes, and knowledge involved in examples of mathematical thinking Interviewing procedures here are typically prescribed, since specifying the details of behavior requires careful channeling of the subject's activities into particular mathematical tasks The interview still varies according to the subject s mathematical behavior, but the primary interest is in competence rather than performance Subjectively equivalent tasks are used, together with flexible presentation and interrogation This insures that the interviewer obtains a good assessment of what the child *can do*, rather than a less useful assessment which may reflect the confusions and misunderstandings arising out of specific phrasings of the problems

While clinical interviewing involves a heterogeneous collection of techniques, there is a common dependence upon the verbal reflections of the subject, on contingent questioning, and on the creativity of the individual interviewer. These techniques raise many controversial methodological questions, since reflection, contingency, and creativity are unorthodox from the perspective of the standardized, objective, replicable scientific method Do subjects' verbal reflections provide accurate data with respect to underlying cognitive processes? In what sense are two clinical interviews comparable? Can two researchers employing clinical interview methods expect to obtain similar results, or, for that matter, comparable results? How can the clinical interview method be used in educational diagnosis? We now consider the issues of the reliability and validity of clinical interview data, and the question of how the clinical interview technique can be used in educational practice

The validity of verbal reflections

TRANSPARENT MIND

Central to the clinical interview method is its dependence on subjects' verbal reflections But reliance on such data is thought by many to be methodologically problematic, if not outright illegitimate. Prejudice against the use of verbal reports is not new Debates about their significance have been part and parcel of the classic controversy over the role of introspection in psychological inquiry. And, as the whole idea of introspective psychology progressively fell into disrepute, the status and importance of verbal data followed suit.

Underlying most early introspectionist claims were certain assumptions about the nature of mind, as well as certain assumptions about people's abilities to report on their mental life The two were connected in that the treatment of verbal reports followed directly from the view of mind presupposed. According to this view, mental life could be identified with conscious awareness; mental states were conscious states The type and content of a mental state was determined by the characteristics of the experience the person had, and this was naturally something the person could be said to be aware of Given this picture of mind and the mental, the rationale for using verbal reports was straightforward Minds are transparent to their owners To have or to be in a mental state is to be in an experiential state to which the mind has direct access So, once we have learned an adequate vocabulary, there should be relatively little difficulty in reporting on our mental life Hence verbal reflections are central data for psychological research

By the early part of the twentieth century, however, this intuitively plausible conception of mind came under serious attack from several quarters. Employing different arguments, logical positivists, behaviorists, philosophers of mind, and Freudians, all challenged the transparent model of mind and in turn cast doubt on the validity of introspective data for psychological research Subsequently, of course, these various programs themselves ran into problems Nevertheless, the criticisms they raised had force. The once appealing conception of the conscious content of mental life could not sustain serious scrutiny and has now been largely abandoned. [A fuller account is given in Ginsburg, Kossan, Schwartz and Swanson, 1982.] Thus, the contemporary cognitivist who wishes to employ the clinical interview method can no longer depend on the introspectionist model of mind to provide a simple rationale for the use of verbal reflections. Moreover, once it is admitted that the mind is not transparent, that mental states cannot in general be identified with states of conscious awareness, and that language is not essentially geared to describing private mental states, questions concerning the reliability and significance of introspective reports begin to loom large. So the dilemma of the enthusiastic clinical interviewer is how to square methodological needs with a reasonably balanced view of mind and the mental We believe this can be done without presupposing various discredited features of the old introspectionist model. Put simply, our position is that: 1) Verbal data do have a place in cognitive research; 2) there are important limits and constraints on their use; 3) effective use of verbal data requires paying careful attention to these limits and constraints; 4) provided this is done, any of the remaining qualms about using verbal reflections are also those which apply to other sorts of data collected by more standard research methods.

SOME USES OF VERBAL REPORTS

There are several areas, relevant to the study of mathematical competence, where there seems to be little reason for *overall* skepticism about introspective reports. One is subjects' claims about their own thoughts and beliefs. True, we can no longer assume, as earlier theorists often did, that such reports are the result of direct intuitions of conscious happenings, or that they cannot possibly be mistaken. And, true, we may have to allow that some of our beliefs are not known to us. But with all this said, the fact remains that over a wide range of conditions and situations people are reasonably good at telling what they believe and think. This seems especially to be the case when it comes to mathematical knowledge. Indeed, with many of our more abstract or complex mathematical thoughts (e.g. "$\sqrt{(64)} > \sqrt{(49)}$" or "The area of a circle equals πr^2"), it would be difficult to make sense of the claim that a subject had such knowledge independent of an accompanying ability to articulate it in a language or other symbol system. So there is good reason to believe that the clinical interview can be a useful tool for securing information about the facts and principles subjects may employ in their mathematical reasoning.

Second, we think it is a reasonable practice to rely on subjects' descriptions of *some* of their cognitive processing. This in part depends on how one construes the notion of *processing*. Construed narrowly enough — e.g. limited to descriptions of neural processing — introspective reports will be of no help. On the other hand, if stages or steps taken in solving a problem are the aspects of processing that are of interest, verbal reports may be a valuable source of information. For example, in arithmetic there is no reason to distrust subjects' claims that they did a column addition problem starting at the bottom, say, or in logic that they began with the conclusion and worked back. And while it may be admitted that a person has little or no access to some of the sources of his insight or to the underlying processes responsible for creative synthesis, even here all needn't be opaque. The mathematician may be able to recall approaches that failed or key realizations that turned the problem around. Perhaps with the aid of drafts of earlier proofs he may even be able to provide us with a detailed understanding of why one assumption was dropped, another changed, and a new strategy adopted at this particular point.

Therefore we believe that introspective reports can provide useful information about mathematical knowledge and processing, and that the clinical interview method has a place in such research. Giving introspective reports credence, however, does not mean that we buy the old introspectionist model of mind. Introspective reports are not in general based on direct "inner perceptions" of conscious qualities; such reports may be in error and other tests may have to be used to confirm or refute them. But the clinical method itself has resources that are available to help check on subject error. The interviewer can always pose additional questions, challenges, or tasks, to test the soundness of any suspect report.

RELEVANT DOMAINS

Introspective reports of mental states and processes can be expected to be accurate only within domains to which subjects have access. Now it should be clear that subjects are frequently not in a position to know about various areas of their cognitive activity. Even the most committed old-time introspectionist allowed that the mechanisms underlying sensory and motor skills were not all available for introspective report.

How informative subjects' reports will be therefore depends on what sorts of activities they are questioned about. It would be most helpful, of course, for investigators to have at hand a general model of mind and cognition that would tell in advance which processes and states their subjects have access to, which they don't, and why. But it is unlikely anything like this is going to be worked out in detail prior to developing better theories and understanding of the individual domains themselves. So the whole thing must be something of a bootstrap operation. Nevertheless, as we have argued in the previous section, it seems to be an empirical fact that in some areas of mathematical thinking, subjects can report on various aspects of their activities. Equally, a certain amount of common sense and exploratory inquiry indicates that in the case, say, of reading, it would be foolish to suppose that subjects can report on deep principles of grammatical structure, on how sentence meaning is derived, or where exactly were their points of eye fixation. What information is accessible to subjects will vary from domain to domain.

SELECTIVITY AND LEVELS OF DESCRIPTION

The object in front of you is a pencil, a wooden thing, contains cellulose molecules, is blue, is navy blue, reflects light of predominantly 470 mu wave-length, is smaller than a breadbox, and is not made in Syracuse. The number of correct descriptions is unbounded; hence a *complete* characterization is impossible. Any attempt at description is always constrained by the need to select aspects and levels of reporting. At the same time such selection is constrained by the knowledge and data available to the person doing the describing. We may have the knowledge and information to report on some aspects and not on others and, within an aspect, we may be able to report at one level of detail and analysis and not at another. These same considerations come into play when dealing with verbal reports of mental states and processes. What a subject reports will always involve selectivity and interpretation. Introspective descriptions are not representations of an unconceptualized given, but, of necessity, reflect the subject's skills and habits of categorization.

When we ask a subject what she did or how or why she accomplished a task, we must allow that there is no one correct answer. Within a domain accessible to the subject there will always be alternative levels of analysis, some of which the subject may be aware of and others not. Consider, for example, a subject's attempts to reflect on a mechanical skill like tennis. The subject may describe her last tennis serve sparsely, "I changed my grip this way and threw the ball higher," or she may give more specific details about movements, steps, and adjustments along the way. But, of course, sooner or later, in trying to be more precise, she will hit a level or type of characterization — be it neurological or at a finer specification of eye, finger and arm movements —

to which she has no access and cannot accurately report. The same holds for descriptions of mathematical activities. In reporting on a long division problem worked mentally, a subject may merely indicate that she figured out that the initial number was 4 and then decided that 25 should be the next number. Alternatively, she may explain in detail how she grouped the numbers, and in what order she applied subtraction, multiplication and simple division. She may still have no idea of the processes underlying her choice of strategies, and may be able to say no more about how she multiplied 9×5 than that she knew it equaled 45.

So in seeking to acquire fruitful data on states or processes via the clinical method, care must be taken to assure that the subject has access to descriptions at the level of analysis that interests the investigator. And where various levels are accessible, steps must be taken to secure reports at the level relevant to the researcher's project. For this purpose, however, the clinical method need not rely solely on context, task structure, or common conversational gambits to lead the subject to the type and level of report desired. The clinical interviewer can take a more direct approach by specifically indicating the level and by explicitly questioning subjects about those aspects of interest. In any case, it must still be recognized that self-reports can never be "complete" or reveal "all" the causal or processing factors of interest. But, then again, no source of data is ever complete or can provide answers to all of our questions.

REPORT INTERFERENCE

That reporting on mental states and processes might interfere with or change the very nature of the mental phenomena was an issue much debated in earlier arguments over introspection. Many theorists who accepted the transparent model of mind nevertheless doubted that accurate reports could be given of *on-going* mental activity. For example, Brentano argued that were we to stand back and as calm observers dispassionately focus attention on our anger, we would no longer be in that state. Similar doubts about the interference effects of verbal reporting have been raised against the clinical interview method. And surely with regard to various states and tasks the doubts are justified. It's a commonplace that if we attempt to focus on how we are doing some kinds of things, we may not only affect the course of the process, but may no longer be able to complete the task. And even when we succeed, adopting a reflective attitude may affect such variables as reaction time. There is also the possibility that calling attention to the need for reports may make subjects self-conscious and lead them to employ different strategies or means than they might if left on their own. As earlier theorists maintained, under some circumstances it may be more fruitful to ask for reports only after the task is completed. Again there will always be the danger of memory decay, interference, and the subject's merely making up an account of what she thought probably went on. And there is no way to eliminate the influences of past experience, expectation, and habits of categorization and organization. All descriptions, by necessity, are filtered through the reporting subject's conceptual schemes and systems of interpretation. Such reports are never certain, and in doubtful cases it would be wise to test their correctness by other means. But once we allow for the possibility of error, there seems to be no reason to reject all process reports out of hand. The fact remains that some reports provide useful information and do not appear to be seriously distorted by the factors described above.

AMBIGUITY

Ambiguity is another source of variation in verbal reports and its effects must be taken into account when analyzing protocol data. It may arise in subjects' responses to "why-questioning." We think it important to distinguish at least three types of answers that may be run together in the protocols. We label these different categories: 1) processes 2) explications and 3) rationales. This classification scheme is neither exhaustive nor precise, but we believe the sorts of distinctions it points to are of particular relevance to studies in mathematical cognition.

Suppose a subject is asked to describe how he did a column addition problem and he reports that he added the column on the right and carried a 3 to the next column. If asked why he did this he may reply that the first column came to 32 and that he was taught a procedure that requires putting the 3 in the next column. He may even be able to spell out the procedure in detail. These answers then represent the subject's attempt to report on his processing activity. Alternatively, the why-question might be construed as asking for the subject's explanation of the carrying maneuver, in the sense of "Why does carrying give the correct sum?" Were the subject to understand the question in this manner, he might cite facts about addition or the notational system. Both sets of verbal reports may be accurate but they are answers to different questions. Notice, too, that the subject may be able to add skillfully but not know why the procedure works, or he may be able to provide an explication but have little skill at actually doing computations (e.g. due to memory or attention failure).

A related, but somewhat different way in which why-questions may be answered is by the subject offering what we might term *rationales*. Suppose a subject is asked why he called a mark an instance of the numeral 4. The subject may, on the spot, make up an explanation referring to its particular angles, its connecting lines, and contrast it with other nearby numerals and letters. In cases like this we know that people are usually unable to provide definitions or defining properties that determine the application of concepts like '4-shaped', but asked to say something they will pick salient features of the given instance. Furthermore, if the subject could spell out a set of features that characterized all and only 4s, it wouldn't follow that such an analytic scheme played a role in the actual processing or provided an account of why the procedure actually worked. After all, we can define a triangle as a polygon whose angles sum to 180°, although it is unlikely that angle measurement and summation precede our determining to label something a triangle. Rationales in protocols may thus provide interesting information about the knowledge and beliefs of the subject, but are not accurate as reports of mental processing. And while they may be offered by the subject to defend or support, and in that sense justify, some action, they are not to be interpreted as providing an explanatory analysis of the particular processes or steps actually taken.

Potential ambiguity of subjects' responses is thus a feature of the clinical interview, but the directive nature of such interviewing allows the experimenter to specify explicitly the type of report required or to probe for clarification when the protocols seem ambiguous.

INDIRECT REPORTS

Earlier we outlined what might be called the "clinical interviewer's dilemma." Breaking with behaviorism and S-R psychology, the cognitivist wants to talk about internal states and mechanisms. But abandoning the introspectionist model of mind means that there is no assurance that the subject has access to the features of cognition the theorist is interested in mapping out. In addition, the deeper and more abstract the structures or processes under consideration, the less likely it is that subjects will have access to them. Thus there are serious limits and constraints on how far it is possible to go with verbal reports of this direct nature. On the other hand, we believe there is no reason to assume that verbal reflections can be useful only in cases where the states or processes under investigation are accessible to the subject. Data gleaned from verbal reports may be valuable in generating and testing hypotheses when the reports are not *directly* linked to the particular hypothesis of interest.

When Piaget and others ask subjects to report on justifications or rationales for a decision or answer, they do it not merely to catalog justificatory practices but to gain insight into the concepts and structures the subject tends to employ in the domain. For example, the subject's particular justification of a conservation judgment is not important in and of itself. Rather it serves as data for more theoretical hypotheses about the types of logical reasoning a child can or cannot engage in at a certain stage of development. But these features of cognition are not themselves subject to direct reports of the child. Another example of an even more indirect use of verbal data, albeit somewhat removed from the mathematical domain, can be found in the work of Freud. Although Freud challenged the introspectionists' mind-consciousness equation, his method of inquiry — in overall theorizing, as well as in dealing with individual patients — depended heavily on protocol data. Not only does Freud allow that verbal reports may be inaccurate, it is part of his theory that important mental states may be inaccessible to the subject. What's more, the theory predicts that certain verbal reports will be mistaken. It is the essence of a repressed belief or desire that the subject will deny its presence. Freud nevertheless uses introspective data as a fertile source of material for generating and testing hypotheses about mental states, structures, and processes. With Freudian theory, then, we have a case where a verbal report denying some hypothesis H is taken to support the claim that H truly characterizes the subject. Verbal reports, therefore, can provide valuable *indirect* data concerning cognitive states and processes.

DATA AND THEORY

The above example highlights a point about the function of evidence in general and the function of verbal data in particular. The use and significance of data is always relative to theory and to those other hypotheses and assumptions held about the domain under consideration. Data serve as evidence only against a set of background beliefs and assumptions. Individual hypotheses are almost never entailed by data alone. The richer and more theoretical a hypothesis, the more complex and tenuous is the route back from hypothesis to observable data. Verbal reports, or data secured by experiments and observations of other kinds, cannot serve to *test* hypotheses in isolation. When it comes to *generating* hypotheses, the link between evidence and theory is even looser. At most, data can be suggestive or point the way. Protocols, by themselves, are neither invariably fruitful nor invariably banal; it all depends on what we do with them. This, of course, means that the value or significance of any set of data will be a function of the intelligence and perspicacity of the theorist using it. To evaluate the fruitfulness of verbal data is to look to see what its payoff has been or is likely to be. And in the case of research on mathematical thinking, we feel the payoff has already been significant.

Contingency

THE RESEARCH CYCLE

A defining characteristic of clinical interview methodology is its contingent structure. The specific direction an interview takes — the questions that are asked — varies as a function of the subject and the subject's answers to earlier questions. Such variability raises questions of reliability and validity. To what extent can clinical interviews be replicated? Do clinical interviews yield valid data? The answers to these questions are complex because research, including that using the clinical interview method, is characterized by a cycle of investigative activity. The cycle varies in its investigative purpose and in its theoretical specificity. During initial work in a new domain, the primary function of research activity is to discover. In the study of mathematical thinking, initial work focuses on the identification of interesting phenomena and the development of useful categories of mathematical knowledge and cognitive processes. After a period of discovery, an identification or theory specification stage ensues. In this stage of the research, theoretical ideas on cognitive structures, knowledge, and the processes involved in solving mathematical problems become more developed. As successful theory is generated and tested on one level of detail, new theory is needed at the next, more detailed level. Thus, theory and research evolve cyclically, as more and more detailed understanding of mathematical thinking is achieved.

Regardless of the level of theoretical specificity, any individual study may be oriented toward the purposes of hypothesis generation or hypothesis testing. During the discovery stage, initial activities are likely to fall in the hypothesis generation category, since very little is known of the research domain. After some initial investigation, however, during which some very general hypotheses are identified, a theory-testing phase of research is begun. A typical hypothesis during the discovery stage might be as broad as, "Children count on their fingers when first learning to add." Research efforts then collect data bearing on this hypothesis. During the specification stage, research can still be oriented toward hypothesis generation or hypothesis testing, depending on whether the aim is the elaboration of existing theory or the testing of theory already constructed. Thus having determined that children learning to add often count on their fingers, a researcher might then attempt to construct a more detailed theory of the strategy employed, including the role of "counting on." Initial efforts toward this goal would involve hypothesis generation — the construction of a hypothesized strategy for beginning adders. Following hypothesis generation, which increases the level of theoretical detail (as compared with the discovery stage of research), the researcher then attempts to test the newlygenerated theory. Thus research can be characterized as involving hypothesis generation and testing at both the discovery and theory specification stages. We consider now the issues of reliability and validity at different stages of the research cycle.

HYPOTHESIS GENERATION

The hypothesis generating stage of research is characterized by a lack of theoretical development at the level of the research interest: more precise characterization of mathematical thinking is desired. A major criterion for the use of a particular methodology during this stage is its utility in fueling the investigator's intuition, in uncovering new phenomena, and in providing rich descriptions from which to induce theory. Under these circumstances it is difficult to argue that any particular methodology is best. Virtually any technique useful to the individual investigator is acceptable. In practice the clinical interview is highly suitable, since its flexibility permits exploration, and since it provides the subject's own view of the knowledge and cognitive processes involved. As we saw above, while the subject's view of his own cognition cannot be taken as infallibly correct, it surely provides relevant information for hypothesis generation.

Reliability and validity of measurement are not relevant issues during the hypothesis generating stage of research. These psychometric issues become important only when the specific purpose of measurement has been determined — when the constructs to be measured are theoretically specified. By definition, during the hypothesis generating stage of research there is a lack of theoretical development: the purpose is to extend theory to a new level. If research is aimed at the discovery of interesting phenomena, the criticism that interviewers behave differently or obtain different results is not appropriate. Similarly, if research is aimed at hypothesis generation at a more detailed level of theory, some initial difficulties in construct definition and theory development are to be expected; psychometric precision is simply not a relevant issue until such development has taken place. Consistency of measurement (reliability) must be unimportant if the purpose and processes of measurement are still being worked out. The accuracy of the inferences (validity) drawn from the data cannot be a serious problem if the types of inferences to be made have not been identified. Thus during the hypothesis generating stage of research, psychometric criticisms of the clinical interview are not appropriate.

The hypothesis generating stage of research, in which the flexibility of the clinical interview is a great advantage and in which psychometric considerations are not relevant, is usually dismissed as exploratory and preliminary. The implication is that such fooling around is secondary to the truly "scientific" job of testing hypotheses in a rigorous manner. This view is but a caricature of science. In our achievement of understanding there can be no testing of interesting hypotheses or insightful ideas without first generating them.

HYPOTHESIS TESTING

When theory is available and the predominant research issue is hypothesis testing, the justification and rationale for the clinical interview method must address psychometric issues. In this situation the clinical interview can no longer be viewed as a *research* method — that is, a method for discovering phenomena and creating ideas; rather, the clinical interview must now be considered as a *measurement* method — that is, a way of acquiring data for theory testing. Now psychometric requirements are of major concern.

Idiographic barriers to hypothesis testing. In the area of mathematical thinking, idiographic and historical factors pose serious barriers to hypothesis testing. Current theories typically describe mathematical thinking in idealized and general terms, and are concerned with competence over a range of problems rather than with idiosyncratic performance on specific problems. Theoretical notions of competence must, however, make contact with performance factors. Any particular subject, prior to participation in a study, has had extensive mathematical experiences, often specifically in the area under study. "Sesame Street", teaching by parents, the school curriculum, games and books, and spontaneous learning may all contribute to an individual's knowledge. Thus the subject's history may obfuscate hypothesis testing; the link between the idealized theory and the realization in which it is empirically tested is often quite tenuous, since the subject's history usually varies idiographically. The interview with Patty [in Ginsburg, 1981] provides a nice example of idiographic barriers to theory testing which are surmounted through contingent interviewing procedures. Patty's contact with addition prior to the interview had somehow resulted in a differentiation of the mathematical meanings of "altogether" and "plus." "Altogether" may have evolved its meaning for her through informal, everyday contact with addition, while she seems to define "plus" in terms of school algorithms. The flexibility inherent in the clinical interview provided a means of deciphering these idiosyncratic aspects of her mathematical learning.

Theory provides general notions for analyzing the cognitive structures, knowledge, and processes involved in an individual's mathematical thinking. But contingent methods are required to work out the application of the general theory to the particular case. Without contingent procedures it is very difficult to come to terms with the complexity of idiographic behavior and to achieve healthy theory development and testing.

Psychometric considerations. When clinical interviewing is used for hypothesis testing, no unusual psychometric difficulties need arise. For example, consider a hypothetical study using the clinical interview method. Before the subjects are tested the interviewing procedures have been at least broadly defined. Thus the interviewer generally knows what types of questions he will ask and which materials he will use. Further, he has an explicit set of subject attributes for which he seeks "scores." "Scores" can be interpreted quite broadly; e.g. a score might be the assignment of a subject to a category like "counts on from the larger number", or "uses finger counting". Each *subject* is interviewed by several *interviewers*. Each interview involves several *problems*. Each problem yields a set of *behaviors* that can be categorized. The interviews are then transcribed and several *raters* read each transcript and determine the *scores* on each problem in each interview.

Traditional conceptions of reliability and validity apply in this situation in the usual way. Consider, for example, inter-rater reliability. For each combination of subject, problem, and interviewer, several raters have scored the interview From these data, the overall agreement among raters' categorizations can be determined; this is an estimate of inter-rater reliability Inter-interviewer reliability can be calculated in a similar way Several interviewers carried out interviews with each subject. The agreement among these interviewers can be determined; this is an estimate of inter-interviewer reliability. By viewing problems as "items", estimates of internal consistency (homogeneity) can also be derived. Thus one can determine whether subjects receive similar "scores" on similar items Generalizability theory [Cronbach, Gleser, Nanja and Rajaratnam, 1972] provides a single framework for conceptualizing the various reliability coefficients in this situation just as it does in more traditional (non-contingent) situations. In brief, in clinical interviewing, as in standard testing, one can determine whether raters agree in their scoring of data, whether independent testers obtain the same data from the same subjects, and whether subjects' behavior is consistent within the bounds of the interview.

Several comments need to be made concerning the different aspects of reliability First, inter-rater reliability usually reaches acceptable levels in systematically-conducted clinical interview studies Thus raters usually have little difficulty in agreeing on the scoring of clinical interview transcripts [see, for example, the reliability figures in Houlihan and Ginsburg, 1981] Second, in mathematical problem solving situations, one does not necessarily expect high levels of internal consistency on the part of the subjects. The child may employ one strategy on one addition problem and another strategy on a second. For example, Resnick [1976] found that young children used different subtraction strategies depending on key characteristics of the problems. Ginsburg, Posner and Russell [1981] have obtained similar results in the case of mental addition Internal consistency, therefore, is not always to be expected. Third, inter-interviewer reliability — agreement between two independent interviewers of the same child — is a complex matter deserving special attention So far as we know, there have been no empirical investigations of this type of reliability For technical reasons such investigations are difficult to conduct, since the first interview may "spoil" the child — that is, bias his future responses, teach him, etc. Even if it were possible to conduct such research, low reliability might still be obtained. For example, interviewers can differ in their interviewing skills and judgments This situation, however, does not necessarily reflect on the interview method itself; the solution is to use skilled interviewers. While some unanticipated variability in the subjects' behavior in the interviewing situation often does occur, careful preparation of interviewers minimizes the problems that arise. Along a different dimension, two interviewers may obtain different results because they are interested in different problems. This situation, too, does not necessarily reflect on the clinical interview method itself When the interviewers' interests differ, similar procedures leading to different results are perfectly acceptable, and indeed to be expected. No reliability problems arise here — reliability of a *research* method is not psychometrically defined; only reliability of a *measurement* method is.

Traditional conceptions of validity also apply to the hypothetical research study outlined above The accuracy of the inferences drawn from the data collection/analysis procedures can be investigated in normal ways For example, patterns of the relationships with other measures of interest (arithmetic achievement on standardized scores, class performance on problems similar to those used in the study, etc) can be established as dictated by the theoretical substance of the study. Oftentimes, because verbal protocols can be supplemented by observations in the interview settings, by eye-movement data, and by latencies, or other process data, rich validation information is available.

Throughout the preceding discussion of reliability and validity it has been supposed that the interviewing procedures are predetermined, at least to some extent It has been assumed that appropriate relationships among contingent interviewing procedures, subject behavior, and scoring of that behavior, have been worked out in advance. Yet very few investigators using the clinical method specify *a priori* the contingencies used in interviewing — they simply "go with the flow" of the interview, leaving implicit the adaptation of interviewer procedures to subject responses. Analytical procedures may similarly be related implicitly to the theoretical constructs, with interview "scores" determined "holistically." This approach may or may not present a problem, depending upon the skill of the interviewer, the complexity of the inferences that are made, the subject matter area, and so on. Hypothesis testing in the discovery stage might proceed quite successfully in this way, since the theory to be tested is fairly global — e.g judgments on the use of finger-counting are fairly straightforward Research addressed to specification is more likely to be adversely affected since the necessary judgments are more subtle When interrelations among contingent questioning, subject behavior and scoring are unspecified, the reliability and validity of the interview as a measurement tool will probably be adversely affected This is simply poor *use* of the methodology, not an indictment of the method itself

DIAGNOSTIC TESTING

Several applications of clinical interviewing to education are possible. See for example Trivett's [1981] comments on the use of clinical interviewing in teaching Here we discuss the role of clinical interviewing in educational diagnosis Clinical interview techniques suggest new directions for diagnostic testing Instead of employing standard tests as, for example, "Key Math", the diagnostician can operate as a clinical interviewer. She can present a set of (sometimes standardized) problems to a child, engage in flexible and contingent questioning, and generally intervene in any way that will produce further information concerning the child's knowledge and problem solving processes. In such diagnosis by clinical interview, the examiner develops hypotheses concerning the child's difficulties and strengths and tests these hypotheses, as in other research efforts In this view, the diagnosis of mathematics difficulties is research in miniature The process of diagnosis resembles theory development and testing in research on mathematical thinking, but is always focused on the individual case. The diagnostician is a researcher: she generates and tests theory. Like research, diagnosis involves discovery, specification, and the determination of competence.

The initial phase of diagnostic testing involves discovery The diagnostician — who may be a school psychologist, a teacher, a clinical psychologist, a pediatrician — usually has some vague information to the effect that the child under consideration has been performing poorly in mathematics. The initial information is usually of a general sort; for example, that the child has received low grades in school, or a failing score on an achievement test. Under the circumstances, the diagnostician's first task is broadly to identify the child's weaknesses, and also his strengths. Does the child have a concept of place value? Can he carry? Does he employ a particular error strategy in calculating subtractions? The diagnostician formulates hypotheses concerning the child's difficulties. Each hypothesis is in effect a miniature cognitive theory of the individual child. If at all sophisticated, the hypothesis usually involves considerating a configuration of cognitive processes: that the child employs a particular error strategy, and also has a particular type of knowledge concerning place value, and also misunderstands the textbook presentation of the concept in question. Thus discovery involves the uncovering of interesting phenomena — e.g., the child adds by using the numbers on the face of a clock — and the generation of interesting hypotheses. Usually the hypotheses that are generated revolve around the child's problems — his failure to calculate properly or his misconceptions. Less frequently the diagnosis also generates hypotheses concerning the child's intellectual strengths; for example, his informal and untaught mental calculation procedures. In our view this aspect is also essential since most mathematics learning difficulties stem not from some deep-seated psychological problems of the child but from inadequate instruction in the schools [see Ginsburg and Allardice, 1982, for a discussion of this view]. In any event, once hypotheses concerning difficulties or strengths have been generated, they need to be tested, and this too can occur within the context of the diagnostic clinical interview. If the examiner discovers that the child has an unanticipated difficulty in finger-counting, then tests of this hypothesis are devised and conducted on the spot in a flexible manner. If the examiner discovers, as in the case of Patty, that the problem appears to be linguistic, then appropriate tests of this hypothesis are immediately conducted. In brief, the first stage of diagnosis is discovery, involving the noticing of interesting phenomena, the generation of hypotheses, and testing them. The creative diagnostician functions in the same way as the creative psychological researcher, although the former's aim is a theory of the individual whereas the latter's is a theory of thought across individuals.

The diagnostician's next task is specification. The broadly identified difficulties the child is experiencing, as well as his strengths, must be defined in a more precise manner. Thus, if in the discovery phase the diagnostician suspected some problem with written calculation, in the specification phase the nature of the difficulty must be clearly identified. The distinction between discovery and specification is only one of degree; the two phases often blend one into the other. Again, in the specification phase, the diagnostician employs both hypothesis generation and testing. The intent is to generate precise hypotheses — or at least more precise than in the discovery phase — and to test the hypotheses in a rigorous manner. The specification phase ends when the diagnostician feels that a sufficiently precise "theory" of an individual's mathematical thinking is available. Of course, precision is only relative: after the initial specification, further discovery may be indicated, which in turn leads to even more precise specification. Like research, diagnosis is a cyclic phenomenon in which earlier knowledge is made to appear imprecise and even naive by later developments.

At both the discovery and specification stages of research, the diagnostician must focus on the distinction between competence and performance. By definition, children experiencing difficulties in mathematics exhibit low levels of performance. This much is obvious; the children would not have been referred for diagnosis if it were not already known that their performance was below average. If the diagnostician's only accomplishment is to predict that the child will perform poorly, then diagnosis will have accomplished little indeed. (This is not a far-fetched possibility: most standard tests do little more than this.) To make a genuine contribution to the understanding of the child, the diagnostician needs to determine whether the child really is unable to understand place value, say, or really cannot compute two-digit addition problems, or really has no informal understanding of addition. Such a concern with competence, as opposed to mere performance, is especially important in the case of children experiencing school failure. For these children, failure in school is the norm; the exhibition of competence is generally reserved for out-of-school contexts. The diagnostician, therefore, needs to work against the expectations inculcated by many years of schooling; she needs to make extra efforts to uncover hidden competences. This situation is not very different from basic psychological research focusing on unusual populations. Thus in cross-cultural research the psychological investigator needs to devise tasks which can overcome his cultural bias and ethnocentric tendencies if he is to obtain an accurate picture of the cognitive functioning of the culturally unfamiliar. This often requires non-conventional research techniques [see for example Cole & Scribner, 1974], and when done well often reveals the presence of unsuspected intellectual skills. The same type of argument can be applied to children who are failing in school mathematics, many of whom are also culturally different — poor and/or black — to the white middle-class diagnostician.

We see then that the process of diagnosis strongly resembles that of cognitive psychological research. Diagnosis involves discovery, specification, and a focus on competence. Just as the clinical interview is useful and effective in cognitive research, so is it valuable for diagnosis. Just as standard tests and naturalistic observation are generally not the methods of choice for psychological research into cognitive process, so they are often not the most effective techniques for diagnosis. The clinical interview is a research and measurement tool which can prove useful for individual diagnosis — for the construction of theories about individuals.

References

Cole, M. and Scribner, S. *Culture and thought*. New York: John Wiley, 1974

Cronbach, L., Gleser, G., Nanda, H., Rajaratnam, N. *The Dependability of Behavioral Measurements*. New York: John Wiley and Sons, Inc. 1972

Ginsburg, H. P. The clinical interview in psychological research on mathematical thinking: Aims, rationales, techniques. *For the Learning of Mathematics*, 1981, *1*(3), 4-11

Ginsburg, H. P. and Allardice, B. S. Children's difficulties with mathematics: Cognition in the school context. In: B. Rogoff and J. Lave (eds.)

Everyday cognition Its development in social context Cambridge Mass : Harvard University Press 1982, in press

Ginsburg, H P , Kossan, N E , Schwartz, R and Swanson, D Protocol methods in research on mathematical thinking In: H. P Ginsburg (ed.), *The development of mathematical thinking* New York: Academic Press, 1982, in press

Ginsburg, H P , Posner, J K and Russell, R L The development of mental addition as a function of schooling and culture *Journal of Cross-Cultural Psychology*, 1981 *12*, 163-178

Houlihan D M and Ginsburg, H P. The addition methods of first- and second-grade children *Journal for Research in Mathematics Education* 1981 *12* 95-106

Resnick, L Task analysis in instructional design: Some cases from mathematics In D. Klahr (ed), *Cognition and instruction* Hillsdale N J.: Lawrence Erlbaum Associates, 1976

Trivett, J The clinical interview: A comment *For the Learning of Mathematics* 1981, *2*(1), 51

Acknowledgment

Preparation of this paper was supported by a grant from the National Institute of Education NIE-G-78-0163

TEACHING

Teaching is more difficult than learning We know that, but we rarely think about it. And why is teaching more difficult than learning? Not because the teacher must have a larger store of information and have it always ready. Teaching is more difficult than learning because what teaching calls for is this: to let learn. The real teacher, in fact, lets nothing else be learned than — learning His conduct therefore often produces the impression that we properly learn nothing from him, if by "learning" we now suddenly understand merely the procurement of useful information. The teacher is ahead of his apprentices in this alone that he has still far more to learn than they — he has to learn to let them learn

Martin Heidegger *What is called thinking?* (p 15)

Two Modes of Thinking — also Relevant for the Learning of Mathematics?*

IPKE WACHSMUTH

"2:43 p m — what time is it now, then?" I'll not readily forget my colleague thinking aloud that way when reading off his new digital watch A student told me that, having looked at his old analogue watch, he'd know what time it was; if asked for the time a few minutes later he couldn't answer without having another look first (To get back his sense of time with his new digital watch he first imagined what the hands of his old watch would show)

What these examples could imply in general, and for the learning of mathematics in particular, will be illustrated in the following For the moment we will continue with the above example

What is the difference between the two sorts of watch? The digital watch gives us exact information about the time by means of a *linguistic* string So does the analogue watch but, beyond that, it has a characteristic face, a *"gestalt"*, and displays time through its "facial expression", i.e by the position of its hands Besides communicating the time 2:43 p.m exactly, the analogue watch, by its face, gives us the (pleasing) feeling of having plenty of time yet until "three" (or, about a quarter of an hour later, the alarming one that it must now be high time) In what follows we shall illustrate the hypothesis of different modes of thinking being involved in each case First we will show this more precisely in relation to theories of knowledge representation That such modes of thinking can actually exist will then be shown by some results in modern neurology and neuro-psychology Finally, we will indicate some implications for mathematics learning and teaching, and sketch some projects of recent and current research

*This paper is an extended version of a talk given at the annual meeting of the *Gesellschaft für Didaktik der Mathematik* in Darmstadt W Germany, March 1981

For the Learning of Mathematics **2**, 2 (November 1981)
FLM Publishing Association Montreal Quebec Canada

Knowledge representation

Norman and Rumelhart [1975] propose two extreme possible forms of representating knowledge:

— a propositional system, expressing concepts by means of statements upon conceptual interdependences between the concepts involved;
— an analogous representation, preserving an accurate image of the original scene

Their "one-system hypothesis" is based exclusively on the first possibility which presupposes knowledge to be linked with language; and their model of "active structural networks" is now of great importance in the theory of knowledge representation It disregards, however, an essential component of thinking which could also have a correspondance in representation. The fact that images may be experienced when remembering and processing visual information is explained in their model as an occasional generation of imagery by the propositional system (which is certainly possible: the student who, after reading off his digital watch, imagines the analogous position of hands thereby translates linguistic into visual information) To support their hypothesis Norman and Rumelhart rely on experiments where subjects could "insufficiently" remember images, which they explain, in terms of their model, as arising from conceptual failures in the propositional representation

We consider it possible to oppose their conception that, in such a one-system representation, all human knowledge is assumed to be linked to language The common opinion that man, besides explicit knowledge which he can communicate, disposes of "tacit knowledge", suggests we should allow for the existence of knowledge not linked to language when reflecting on knowledge representation.

In developing a theory of visual perception, David Marr and his co-workers from MIT outline how images and scenes were recognized by the use of "sketches", based only on a general precognition of the scene [e.g. Crick/Marr/Poggio 1980; for further references see this report] In a first step of processing by the visual cortex, the grey-level array given on the retina is transformed *simultaneously* into a "prime-sketch" consisting of lines which in part correspond to the contours of objects, and in part show variations in the surface shape.

Extrapolating Marr's ideas, we could imagine a representation in terms of neural networks adequate to such processes of perception The "core" of such a representation might consist of a storage of still more reduced sketches (like pictograms) where net representations of more complex structural features of the perceived scene are linked We could then assume a more complete image to arise from the activated subnet of the core representation through "resonance", in an act of recognition; much as a stringed instrument, when it is animated by a pure tone resonating with its body (as a "core sketch") produces a sound of some sonority (as a more complete image).

To illustrate this we give an example The possibility of recognizing the picture shown in Figure 1 would accordingly be based on some "analogous" core representation,

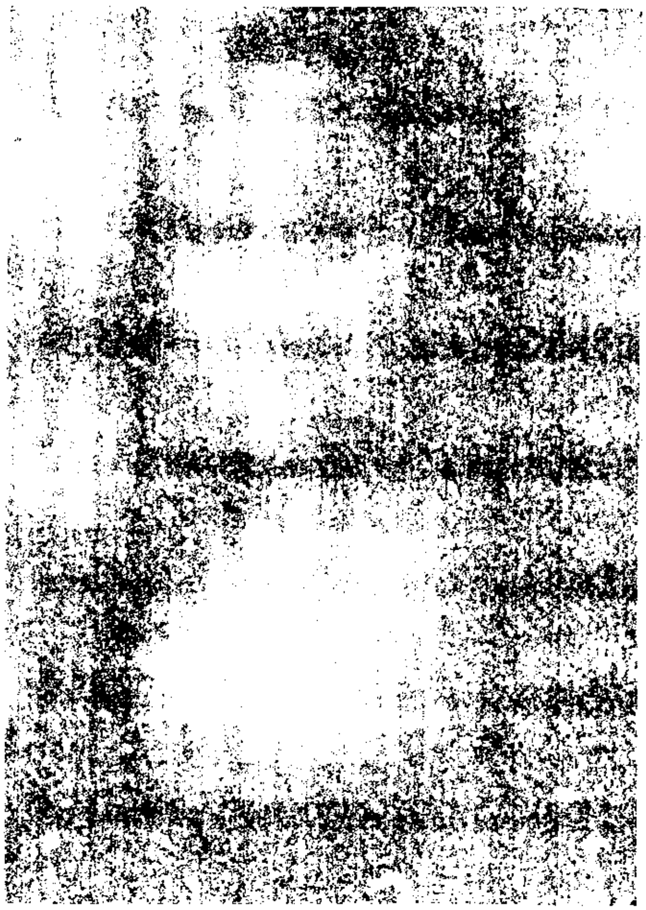

Figure 1