The CGD Project:

An Educational Diagnostic System Based on THEORIST

Marlene Jones

Research Report CS-87-26

April 1987

The CGD Project:

An Educational Diagnostic System Based on THEORIST

Marlene Jones

Dept. of Computer Science,
University of Waterloo,
Waterloo, CANADA

and

John McLeod, Gladene Robertson, James Tubman and Barbara Wasson

Dept. for the Education of Exceptional Children,
University of Saskatchewan,
Saskatoon, CANADA

Abstract

The intent of the Computer-Guided Diagnosis (CGD) project is the development of an expert system to assist the user in diagnosing learning disabilities. We discuss herein the motivation for the project, the application domain including a model of educational diagnosis, as well as previous research and development regarding expert systems within this application domain. We then present specifics regarding the CGD project including the underlying educational and computational philosophies. In particular, we present the details of the implementation of the expert system in THEORIST, a logic programming deduction system developed at the University of Waterloo. We argue that a THEORIST-based expert system for educational diagnosis is particularly appealing because it captures the essence of what expert diagnosticians do, ie. theory formation. Combining THEORIST with our philosophy of employing expertise based on clinical, rather than actuarial experience, promises to produce a system that closely resembles the ideal behaviour of an expert in the performance of an educational diagnosis. In addition to reporting on our initial experimentation with THEORIST, which involved the development of an expert system for diagnosing arithmetic difficulties, we briefly discuss future research and development, as well as current related research projects.

1. Introduction

One major concern of educators is the efficient and effective diagnosis of students' learning problems. In addition to being a major clinical concern, educational diagnosis is an enormous field of research within both psychology and education. As is the case with many large, complex domains, there is not total agreement among the experts. In particular, causes, characteristics of various groups of learners, and appropriate remediation techniques are among the topics which are, at times, hotly debated in the research and popular literature. However, there is general agreement regarding the importance of educational diagnosis, professional guidelines and procedures, and even models of educational diagnosis.

There are many different reasons why a child may be referred for diagnosis – lack of appropriate achievement, behavioural or emotional problems, *etc* – and numerous areas of exceptionality with which one may be concerned – mentally-handicapped, physically-handicapped, visually-impaired, hearing-impaired, learning disabled, emotionally disturbed, socially disadvantaged, gifted, talented, *etc.* We are particularly concerned with children who are referred to as "learning disabled" ie. those children "which exhibit a disorder in one or more of the basic psychological processes involved in understanding or using spoken or written language. These may be manifested in disorders of listening, thinking, talking, reading, writing, spelling or arithmetic. They include conditions which have been referred to as perceptual handicaps, dyslexia, developmental aphasia, *etc.* They do not include learning problems that are due primarily to visual, hearing, or motor handicaps, to mental retardation, emotional disturbance, or to environmental disadvantage (National Advisory Committee on Handicapped Children, 1968, p4)." (Although this definition was first adopted by the United States, it is now widely accepted elsewhere.) In other words, the learning disabled child can be thought of as a child of average or better intelligence, who despite the best efforts of both the child and the teacher (and the absence of the aforementioned handicapping conditions), is not learning within the regular classroom environment.

Throughout North America there are professional guidelines and procedures as well as legislation governing the process of educational diagnosis. Within Canada, guidelines regarding appropriate procedures, differentiated teacher qualifications and responsibilities are outlined in the SEECC (Standards for Educators of Exceptional Children in Canada) Model (Committee on Teacher Education and Professional Standards, 1971). Individual provinces also have various pieces of legislation that specify the rights of certain exceptionality groups: for example, the right to an appropriate education up to a certain age, the guarantee of funding, *etc.* In the United States, Law PL94-142 requires educators to classify children with special educational needs (similar to the different areas of exceptionality mentioned above) and to develop individualized educational programs (IEPs) for them.

Throughout the research literature and the professional guidelines, one fact which is stressed is that educational diagnosis should be undertaken as soon as possible. There is evidence that the longer the child's problems go unattended, the more severe they become: the farther the child lags behind his/her chronological peers and the more likely behavioural and emotional problems will compound the situation. Ideally, assessment should take place within the regular school environment by a qualified person who is familiar to the child (such as the classroom teacher or resource room teacher).

Unfortunately, educational diagnosis does not always follow the recommended guidelines. Often there is a lack of qualified personnel (such as educational psychologists or resource room teachers) or such personnel feel insecure in carrying out the educational diagnosis despite the fact they have the appropriate educational training (this is often the case with resource room teachers) or no initial screening is undertaken before a referral is made to the school psychologist or an outside agency. There is a general trend throughout North America to improve teachers' general training with regard to both the assessment and the instruction of exceptional children. As more teachers receive such training, the initial diagnosis (with some guidance) could be carried out by the regular classroom teacher or the resource room teacher. This is not an attempt to remove the school psychologist from the process – far from it. School psychologists tend to suffer from extremely heavy case loads. If more initial screening is undertaken before the psychologist sees the child and if less complex or dramatic cases are handled by the resource room teacher, the result should be better utilization of the psychologist's time and expertise as well as the efficient handling of many more cases.

It was a desire to facilitate educational diagnosis within the regular school environment that originally motivated the research discussed herein. The main goal of the Computer Guided Diagnosis (CGD) project – a joint effort between researchers at the University of Saskatchewan and the University of Waterloo – is to develop an expert system to guide the user step-by-step through the process of educational diagnosis, from the initial suspicion that there is a problem through to the development of an appropriate remedial program for the case at hand. Although an expert system is a means of providing expertise to the less qualified or less confident user, it is also a tool for experts. It is impossible for a diagnostician (such as a psychologist) to keep up-to-date regarding all research, development and legislation that impacts his/her decision-making. Specific examples include: legislation regarding IEPs, what they must specify, by whom, in what amount of detail, for what purposes; funding formulae, including which test scores are allowable; legislation covering placement decisions; specifics regarding commercially-available standardized tests for educational diagnosis, including reviews of such tests, as well as data regarding reliability and validity co-efficients. An expert system is one means of helping the resident expert to stay abreast of recent developments. It can also serve as a reminder of others' philosophies or theories regarding diagnosis or remediation, or a particular aspect of a skill (such as reading) which one has tended to ignore. Educational diagnosis is a large domain; school psychologists are generally required to be knowledgeable in all aspects of diagnosis at least to the extent that they can undertake the initial diagnosis regardless of the child's problems or handicaps and based on this information make the necessary referrals if another expert's opinion is required. In other words, a diagnostician is expected to possess incredible breadth in the area of educational diagnosis as well as depth. Yet the truth of the matter is that not all psychologists are experts in all aspects of educational diagnosis, nor may it be a sensible use of their time to develop such expertise. Just as the expert system can provide the necessary guidance to a less experienced user, an expert system is one means of supplementing the psychologist's skills and areas of expertise. There are times that any expert need his/her skills refreshed or expanded; an expert system is one appropriate means of doing such. This appears to be particularly true in educational diagnosis where due to time pressures it is not uncommon to become stuck in a routine – administering the same diagnostic instruments despite the fact they may not be the most appropriate for the case at hand. Discovering which are the most appropriate tends to be time-intensive labour, as there are many different factors (concerning the test, the skill and the child) that should be taken into account. With access to relevant test data (stored in an associated database), an expert system can make suitable recommendations regarding test selection. This is also true regarding remediation methodologies and activities.

There are other, although perhaps less major, benefits of an expert system which should be noted. Throughout an educational diagnosis, one wants to detect error patterns; for example, the tendency to substitute whole words when reading, or substitution of the letter 'b' for 'd' at the beginning of words. Most diagnosticians detect these patterns by "eyeballing" the child's work, not by mathematical analysis. Unfortunately, this is one aspect of diagnosis that could afford to be done more thoroughly and more conscientiously. It is, however, one aspect that can easily be relegated to an expert system. Many a diagnostician has complained that precious time is spent writing diagnostic reports; any mechanism that automates at least part of this process is most welcome. Although this is by no means a main motivation of developing expert systems, it is a side benefit.

Finally, there are instructional benefits of both the expert system itself and its development. As mentioned earlier, one reason that resource room teachers are often hesitant to undertake an assessment is lack of confidence and experience. In many cases, their training is limited to one or two diagnostic cases undertaken as part of their university education. An expert system is one means of providing additional training without the necessity of additional children.

But there are other educational benefits that are experienced in the development of the expert system. Firstly, it forces developers to examine the research and clinical literature to establish evidence for much of what is done in practice. As a simple illustration, consider the use of standardized tests. Certain tests have become very popular over the years; however, for some such tests, examination of the reliability and validity coefficients, norming information, and professional reviews has indicated that such tests may not, in fact, be the most appropriate. Because expert system developers are concerned with justifying the encoded expertise, the philosophy of "I'll use test X because I'm comfortable with it and I believe it is a good test" just does not suffice. Hopefully, insisting on more solid justification will improve the quality of educational diagnosis.

The second main benefit experienced during development is obtained from the exact specification of what diagnosticians do during the course of a diagnosis and why. Such knowledge, although difficult to elicit, is invaluable. It is important to remember that eliciting knowledge is not a one-shot effort. Because diagnosticians are often unaware of all the factors which they consider or how they weigh them, the initial elicitation and formulation stages should be followed by a process in which the initial knowledge is refined based on further clinical experience, usually in the form of experimentation.

Hopefully, a long-range benefit of expert system development is determining what factors should be considered. In addition to elicitating experts' knowledge, refining it through test cases and field studies, studies involving the tracking of actual diagnostic cases should be undertaken in order to evaluate the effectiveness of the diagnosis and remedial methodologies. The discussion of such research is beyond the scope of this paper.

The Application Domain

Before turning our attention to previous research and development regarding diagnostic systems for education, let us first present the fundamentals of the application domain. We outline here the basic procedures and assessments involved within an educational diagnosis. The procedure is generally initiated by the classroom teacher who is usually the first to suspect that the child is experiencing some learning difficulties. Over the years, several different models of educational diagnosis have been proposed; for example, see (Bateman 1965, Bond and Tinker 1973, Chandler and Siegel 1982, Otto and Smith 1980, Salvia and Ysseldyke 1978, Wallace and Larsen 1979). These models do, however, generally agree on several aspects of educational diagnosis, such as the basic stages through which diagnosis should proceed (although the naming of such stages may differ) and the type of information that should be gathered in the process. The fundamental stages of a diagnosis are as follows (McLeod 1982):

Retrospective: relevant data about the child's previous developmental history is reviewed;

Definitive: the existence, or non-existence, of a learning disability is established;

Analytic: surface symptoms are subjected to progressively finer scrutiny;

Prescriptive: corrective or remedial action is initiated.

Diagnosis begins with a suspicion that the child is experiencing learning problems and progresses through to an appropriate prescription; however, it does not stop there. The Prescriptive stage is both the final stage of the initial diagnosis and the first stage of remediation. But more importantly, the latter two stages (Analytic and Prescriptive) are cyclic; there should be continual re-assessment of the child's skills and the remedial program should be tailored accordingly. At each of these four stages, several different types of information should be gathered; this data can generally be classified into three categories:

- basic educational skills in areas such as reading, spelling and arithmetic;

- psycho-educational correlates which include those intellectual, visual, auditory and language skill deficiencies that might be related to learning disabilities;
- non-educational factors which are primarily in the medical, social and developmental sectors.

Classification of appropriate data is not always well defined or may vary from one model to another. For example, the child's emotional status may be considered as either a psycho-educational correlate or as a non-educational factor. It is not important how one chooses to classify such data (personal preference can dictate), but the inclusion of such information within a diagnosis is not open to debate. The various levels and stages of educational diagnosis are represented in Figure 1.

During the course of a diagnosis, several standardized tests are generally employed (such as intelligence tests, achievement tests, diagnostic tests for particular areas or skills, tests of social development, *etc*) Most standardized tests fall into one of two categories: norm-referenced or criterion-referenced. Norm-referenced tests allow one to compare the child with his/her chronological peers, and as such they play an important role during the Definitive stage of a diagnosis. Criterion-referenced tests generally include a broad spectrum of items designed to measure several aspects of a skill; the results are used to determine what components of the skill warrant further assessment or remediation. Hence, criterion-referenced tests tend to be used during the later stages of the Analytic phase. Selecting the most appropriate diagnostic test is not a straightforward task; there are thousands of commercially-available standardized tests. Databases of some of the more common tests do exist; for example, see (Jones and Loken 1985, Ralph 1983).

It is important to note that educational diagnosis does not culminate in classifying or labelling the child. A label such as "learning disabled" or "mentally-handicapped" does little to determine the appropriate prescription, as it does *not* provide the teacher with any guidance regarding how to proceed with the child's education. The person developing the child's remedial program needs precise details regarding the child's strengths and weaknesses (ie. a comparison to his/her chronological peers), relative strengths and weaknesses (ie. a comparison to his/her other skills), as well as descriptions of common error patterns, and information regarding psycho-educational correlates and non-education factors. A label, although useful for funding and placement purposes, does little else.

2. Previous Research and Development

Within the field of education, diagnostic expert systems have not received as much attention as those designed for instruction. Moreover, typical software developed for educational diagnosis or aspects thereof does not warrant classification as an expert system and generally concentrates on one small aspect of diagnosis, such as the diagnosis of a particular skill. These criticisms alone do not make such software unworthy, but many an educator has despaired at the lack of flexibility. Herein we restrict our remarks to a few systems built for various aspects of educational diagnosis. These systems have been selected for a variety of reasons; they are similar goals to our project, to illustrate a particular underlying philosophy which is relevant to our research, or to demonstrate a contrast to our approach.

Let us start by examining a couple of systems attempting to provide expertise throughout the first three stages of the diagnostic process: from the first suspicion that a problem exists through to the beginning of the Prescriptive stage. One of the more labour-intensive efforts is that of M-MAC (McDermott Multidimensional Assessment of Children) (McDermott and Watkins, 1985). M-MAC is designed to classify, rather than diagnose; there are 148 possible classifications in 19 areas such as "intellectual giftedness", "conduct disorder" or "specific learning disability". Based on these classifications, M-MAC defines specific behavioural objectives to be used for IEPs. These IEPs may be developed in reading, math, learning styles and adaptive behaviour. Although M-MAC specifies behavioural objectives (which are undoubtedly sufficient to comply with the requirement of PL94-142), they are not sufficiently detailed to form a basis of a remedial program; they do, however, provide the user with a starting point.

M-MAC's expertise is based upon actuarial, rather than clinical, data. Although collecting such data is an impressive task, its clinical validity is questionable. In employing such data, one is relying on the comparison of the norms of standardized tests. In theory, this appears sound, but in truth too many standardized tests are normed on small populations that are often not representative of the general population with regard to race, sex, geography, socio-economic status, *etc.* Even a test such as the WISC-R (Wechsler Intelligence Scale for Children - Revised), an extremely popular individual intelligence test which is generally viewed as being "well-normed", has sample sizes of only 200 children for each age group and the norms were calculated several years ago. Employing norming information and related data as the system's main knowledge base makes the system rigid in that additional expertise earned through years of clinical experience cannot be incorporated.

Because M-MAC's reasoning is based on the test's norms (rather than the tasks employed within the test or the characteristics of the test), only certain tests can be administered. In other words, there is a fixed battery of tests; data from similar tests will not suffice. A second example of M-MAC's rigidity is the fact that the system is unable to compare test scores where the testing spans across the child's birthday. This is because M-MAC is employing the test's norms in order to determine the child's relative standing to his/her chronological peers and such norms are employing the child's age (calculated in terms of number of years and months) as a basis of comparison. This is a common technique employed by most standardized tests; the means by which it has been employed in M-MAC is unfortunate. For example, M-MAC is unable to compare an intelligence test administered a few days before a child's seventh birthday with an achievement test given a few days after; the intelligence test requires the use of norms for children age 6 years and 11 months, whereas the achievement test requires norms for 7 years, 0 months.

Despite M-MAC's limitations, it is worthy of mention on several accounts. Firstly, M-MAC's intended scope is broad; it does not restrict itself to one area of exceptionality. However, at the same time, the system is shallow, not allowing the depth of diagnosis that is necessary within any one area of assessment. Secondly, although we question the soundness of an approach solely based on actuarial-type data, M-MAC's expertise is far preferable to a system based on the apparent rules of a self-claimed "expert". Unfortunately as expert systems become popular within education, the development of small, poorly-designed, unsound systems is inevitable, just as the last decade witnessed an abundance of undesirable computer-based instructional software.

The second system that we discuss here also attempts to guide the user from the initial suspicion that there is a problem through to the point where sufficient information has been gathered so that a remedial program can be developed. This system, developed at the University of Saskatchewan, is actually a precursor to the CGD research project discussed later herein. The expert system, designed for use by someone with qualifications such as those of a resource room teacher, guides the user through the Retrospective, Definitive and Analytic stages of the diagnosis when assessing elementary-age children referred for reading problems. Unlike M-MAC, this system's main concern is not classification, but rather diagnosis. As is the case with all diagnoses (regardless of the reason for referral), the expert system requests information regarding the child's developmental history, current school behaviour, the teacher's and parent's perspective on the child's skills and current problems. At the Definitive stage the expert system employs available data to establish whether or not there exists learning problems within a variety of areas. Skills within academic areas such as spelling, arithmetic, writing are not subjected to scrutiny past the Definitive stage of the diagnosis. Nor are behavioural, emotional or motor problems. However, the existence of a problem within any of these areas would have been detected at the Definitive stage. During the Analytic stage the child's reading skills (as well as related correlates such as auditory perception) are scrutinized (as deemed necessary). The resulting diagnostic report specifies which skills or subskills are mastered or not, as well as indicating relative strengths and weaknesses for the child. Any common error patterns that the child exhibited are also noted. The diagnostic report is sufficiently detailed that the teacher can then plan a remedial program based on the report; the expert system, however, does not contain the necessary database of remedial methodologies and activities.

The reason for selecting reading as the area of attention is because it is the main cause of diagnostic referrals. Secondly, the purpose of this project was to demonstrate to educators the appropriateness and feasibility of developing expert systems even for aspects of the domain for which there is not total agreement among the experts.

The system's knowledge is based on actual test cases from the files of the Institute for Child Guidance and Development at the University of Saskatchewan. The system was developed as a traditional rule-based expert system, programmed in LISP. The results of this initial feasibility study have sparked substantial interest from educators and have resulted in funding for a more elaborate expert system for educational diagnosis; the state of the current project is discussed in sections 3-4. For further information regarding the development and evaluation of the initial system, the reader should consult (Colbourn 1982, Colbourn and McLeod 1983, 1984); sample test cases are also discussed in (Colbourn 1983).

In the development of diagnostic expert systems, reading has not been the area to receive the most attention; arithmetic has been. There are several reasons for this. Early expert systems have tended to be developed by non-educators; the processes involved in arithmetic are in general better understood by the layperson than those of reading. Moreover, there is better agreement among the experts regarding arithmetic skills and the analysis thereof than those of reading or other language skills.

The work of particular note in this application area includes that of Brown and Burton (1978, 1982) and Yonge and O'Shea (1982). The DEBUGGY system, as well as the interactive version IDEBUGGY, is based upon the belief that a student's arithmetic errors are not random but rather are consistent discrete modifications of the correct arithmetic procedure. The system contains expertise regarding common arithmetic bugs and procedural errors. These known bugs can then be compared to the student's answers; based on these comparisons, the system selects a subset of bugs, each of which explains at least one of the student's wrong answers. The elements of this initial set of bugs are then combined to generate additional hypotheses for the particular student. The system must then eliminate as many of these hypothesized bugs as possible. Several heuristics are employed in this task; for example, remove bugs which are subsumed by other bugs. Eventually, each of the remaining bugs is classified according to how well it explains the student's answers, taking into account the number of predicted correct and incorrect answers as well as the number and type of mispredictions. At the end of this process, the best explanation of the student's errors is selected.

One of the appealing features of the DEBUGGY approach is its adaptability to any given arithmetic test. This flexibility is due to how DEBUGGY develops its skill lattice. Given an arithmetic test, the primitive bugs are applied to the individual items on the test, resulting in a partition of the test items for each bug. This allows one to sort the bugs into equivalence classes based on the partitions produced. With each equivalence class one associates a subskill; these subskills then form a lattice partially ordered by the relationship "gets correct answer on all the same problems and more". (Burton 1982). By this procedure one can develop a skill lattice for any given arithmetic test. By comparing the generated lattice with the maximal lattice (where the test would include all possible questions), one can determine which skills are tested and which are not (which is the case if two skills collapse to one node in the generated lattice).

Since the original publication of the BUGGY model and related systems, several criticisms have been raised, largely concerning the complexity of the model as given for subtraction (Bryant 1986; Yonge and O'Shea 1982). Despite such criticisms, it should be noted that the DEBUGGY and IDEBUGGY systems are attempting to perform a real-world diagnostic task. Given a child's answers on a subtraction test, the diagnostician attempts to develop a theory as to what the child is doing; in other words, the diagnostician suspects the presence of certain bugs. The diagnostician however does have an advantage; often in testing situations, single-column evidence can be exploited because the diagnostician is able to watch the child's progress through a given problem.

Several other researchers have since examined the diagnosis of arithmetic errors – particularly subtraction since much of Brown and Burton's work focussed on this domain – with the intent of developing suitable computer-based diagnostic systems (Bryant 1986; Yonge and O'Shea 1982). Such diagnostic systems, designed for the indepth analysis of a particular skill, would generally be included at the Analytic stage of diagnosis. Yonge and O'Shea developed a simple but effective production system. The procedure of subtraction is specified by a set of production rules. It is important to note that one could alter the "coarseness", and hence the number, of production rules, depending on the amount of detail required in the final diagnosis. The idea is to choose a "grain size" which is sufficiently fine to distinguish between the different types of errors and yet course enough that it "could plausibly have been learned in the context of school mathematics" (Yonge and O'Shea 1982 p171). The specified rules are also intended to be independent; that is each rule "represents an intelligible component of the child's total subtraction ability" (Yonge and O'Shea 1982 p171). This allows analysis of the child's subtraction skill based on the existence or non-existence of these components or subskills. Yonge and O'Shea's results regarding the diagnosis of common subtraction errors are equally convincing as those of Brown and Burton; both approaches are effective. What is really at issue is (1) the underlying philosophy and hence the diagnostic approach and (2) gradation or "grain size". The latter should partially be determined by the intent of the diagnosis and the target population. Ideally, one would like to control the "grain size" based on the case at hand (which is not possible in any of the systems discussed so far). The second issue, that of philosophy, we address later.

3. The Current Project

3.1 Project Description

As mentioned earlier, the intent of the current CGD Project is the development of an expert system to guide the user through the entire diagnostic process. This includes the development of an appropriate remedial program tailored to the case at hand. In other words, the system guides the user through all four stages of educational diagnosis. Although the system contains expertise at all three of the levels indicated in Figure 1, the knowledge regarding non-educational factors (such as medical problems) is restricted to an expert diagnostician's (say a psychologist) knowledge of such problems. In other words, the expert system is not an expert with regard to medical factors, but has the ability to detect when a referral is warranted.

Although such an expert system may ultimately have a variety of users such as the psychologist, resource room teacher, regular classroom teacher, principle, other administrators, researcher, training teacher, parent or student, we are initially concerned with the user who does have some background in special education and educational diagnosis. The current expert system is being developed for use by a resource room teacher or psychologist; in other words, someone who is considered qualified to undertake an educational diagnosis but who would benefit from additional expertise and guidance.

3.2 Our Philosophy: The Education Perspective

Over the years, psychology has been greatly influenced by current fashions in other, maybe more academically respectable and secure, branches of science. Educational psychology, and therefore educational diagnosis, has inevitably been subject to the same constraints. In particular, mental measurement has been given some of the aura of an exact science through its associated trappings of statistical methods.

For example, we know from empirical evidence that the individual measured IQ is subject to an error which can be estimated. If a person's IQ is 104 as measured by a reputable intelligence test, such as the WISC-R, we can be assured with 95% confidence, that the child's "true" IQ is somewhere between 95 and 113, ie. between "low average" and "borderline superior". However, the picture becomes even more blurred as additional data are gathered. For example, if the child is considered to be underachieving, he/she would probably be given a diagnostic reading test. If the estimate of the child's reading ability (say, 97 measured in the same units as the IQ test) was as reliable and valid as that of the WISC-R estimate of intelligence, there would appear to be a discrepancy of 7 points between measured intelligence and measured level of reading achievement. But the difference between the child's IQ and her/his score on the reading test is subject to compounded error. In the present example, we can assert with 95 percent confidence that the discrepancy is 7, plus or minus 13, ie. the child might be overachieving by 6 points or underachieving by 20 points, if "expected" achievement is equated with measured IQ. This example alone illustrates our concern with the M-MAC philosophy of basing diagnostic expertise on actuarial data.

However, there exist other, even less simplistic, examples of the current hazards in psychoeducational diagnosis. For example, it is common practice to classify poor readers or poor spellers into "types", yet it has been demonstrated that the reliability with which students can be categorized is far from reassuring (Brownbridge, 1983). The implication, in the context of the present article is to pose the question: is it worthwhile to "sharpen" an expert system for the purpose of educational diagnosis if the focus of the system's attention is itself considerably blurred?

An alternative to the diagnostic procedure described above is to make an informed guess and then cumulatively improve validity by a process of successive approximation; it is also one which is well suited to computer mediation. Further, it represents a situation where it may be argued that the human diagnostician does not have to conform with externally imposed constraints but is able to revert to intuitive practices, in the knowledge that those practices which are based on insight and grounded in experience, will be improved as his/her own fallible storage of clinical recollection is augmented by what becomes accumulated within computer memory.

The literature is replete with examples of diagnostic rules that were originally heuristically-based, but which have achieved a status in themselves – subject, of course, to the "plus or minus error" factor. An intelligence or achievement quotient of 85 has become a benchmark of substandard performance. This happens to be 1 standard deviation below the population mean of 100 for tests such as the WISC-R which has a standard deviation of 15. But half a century ago, Burt (1937) observed that a teacher should be able to teach a student in the middle of Grade 5, and who would be aged 10.0 years, if that student's achievement were no lower than that of the average student at the beginning of the previous grade, ie. aged 8.5 years. Hence, special educational provisions should be considered for students with "Achievement Quotients" lower than 8.5/10.0, or 85. Ostensibly on the basis of empirical study, but in fact largely by arbitrary decision, Morphett and Washburne (1931) enunciated the North American doctrine that a mental age of about 6.5 years is necessary for a child to benefit from instruction in reading. Kirk (1963) declared that if one is assessing a particular facet of a child's communication skills, then any other skills that are inevitably involved in the test procedure should be at a level at least two years lower than the skill that is being assessed. This list could go on.

To rely on conventional, "set-piece" empirical research design and inferential statistics to develop a valid analytical network of complex diagnostic rules, is an impossible dream and the reality beyond the dream is likely to be submerged beneath a sea of compounded error of measurement. A procedure which is based on the best judgement available, and which is coupled with progressive componential evaluation, correction and improvement, has been followed by good diagnosticians since long before the advent of computer technology. With this philosohpy, expert systems promise a qualitatively superior extension of authentic diagnostic practice.

3.3 Knowledge Sources

The task of eliciting and formulating the necessary expertise for each component of the diagnostic model has, in fact, been underway for several years: first as part of Colbourn's original study and then as an on-going major Canada-wide survey (Jones and McLeod 1985). In addition to the team's experts within diagnosis and remediation, a network of more than ninety diagnosticians has been established in order to elicit information regarding diagnosis as it is practiced across Canada. This set of participating diagnosticians (most of whom are school psychologists) represents all of the provinces. When collecting such data, however, it is important to remember that how educational diagnosis is *practiced* is not necessarily how it *should* be carried out. There are many reasons why this is true: lack of time and expertise, availability of tests, cost of acquiring appropriate tests or remedial materials, rigidity or stagnation of some diagnosticians. The participating diagnosticians are providing input regarding how they as individuals practice educational diagnosis and how their entire team or school board proceeds. Some very enlightening facts regarding the practice of educational diagnosis are coming to light as a result; however, a discussion of the data is really beyond the scope of the current paper and will be reported elsewhere.

In addition to this source of data, a set of hand-picked diagnosticians with whom the team can meet on a more regular basis has been established. Data from this group of approximately fifteen highly-regarded diagnosticians – resource room teachers, psychologists, clinicians, professors – is being collected in two manners: structured interviews and brainstorming sessions.

In addition to information gathered directly from experts, data can also be collected from actual case files. Several school boards across Canada have volunteered access to such files (of course, with all identifying information removed). At the current time, however, such data is not being examined.

3.3 Our Philosophy: The Computer Science Perspective – THEORIST

The majority of expert systems developed to date have been production systems. In fact, we too have employed this paradigm for the development of such systems. However, researchers at the University of Waterloo have recently proposed an alternative to the traditional rule-based paradigm for developing expert systems. The new approach is based on the philosophy of constructing and reasoning with "scientific theories" (Jones and Poole 1985, Poole et. al 1985) The idea is to develop a "scientific theory" (Popper 1959, Quine and Ullian 1978) that explains the case at hand. For example, in medical cases, diagnoses can be viewed as "scientific theories" which can make predictions and hence be tested about the patients' problems.

Recall the distinction between DEBUGGY and Yonge and O'Shea's production system. In retrospect, DEBUGGY can be viewed as proposing several theories – in this case, hypothesized bugs – regarding the child's problems. The best theory is subsequently selected. Is this distinction important, if the end result is the same? If the diagnostic procedure more closely resembles that of the human expert, we feel that the answer is "yes". And in the case of diagnostic expert systems, this certainly appears to be the case, as it is the *theory* of what the problems are that constitutes the diagnosis.

THEORIST, a logic programming system developed at the University of Waterloo, provides us with a tool to develop expert systems based on this philosophy of constructing and manipulating theories. THEORIST attempts to develop *theories* in terms of given *facts* and *hypotheses* or *defaults*; these theories then serve as explanations of the given *observations*. In the medical scenario, the observations would include the patient's symptoms and test results. Rather than monotonically deducing theorems from a fixed logical theory, THEORIST distinguishes facts from hypotheses and attempts to use deduction to construct theories consisting of facts and instances of hypotheses, for which the observations are logical consequences. The theory developed is empirically tested, by verifying whether other predictions based upon this particular theory are indeed valid for the case at hand. In addition to checking the theory based upon its predictions, one must also verify that THEORIST's proposed theory is consistent.

A typical THEORIST knowledge base consists of a collection of logical formulae classified as possible *hypotheses* or *defaults* (Δ), *facts* (F), and *observations* (G). The formulae employed within THEORIST are:

F — Facts — The set of formulae which we know are true in the world we are trying to represent. We assume that these are consistent (the intended interpretation being a model).

There are two main subclasses of facts: knowledge about the domain that includes knowledge about the interaction of hypotheses (these are generally referred to as *rules*) and knowledge regarding the particular case at hand (these are termed *case-facts*). Depending upon the diagnostic domain, there may be several different categories of case-facts and rules. Not all of the facts need be specified initially. Some of them may be acquired during the course of the diagnosis by asking the user.

Δ — Defaults or Hypotheses — This is the set of possible hypotheses that we are prepared to accept in an explanation of the observations. These may have free variables in which case instances of the defaults can be made by giving values to these free variables.

G — the set of observations or goals to be explained.

A theory is a subset of possible hypotheses that are consistent, and imply the observations. More formally, G is *explainable* if there is some subset D of Δ such that

$F \cup D$ logically implies G

$F \cup D$ is consistent.

D should be perceived as a "scientific theory" used to explain G.

To illustrate these different THEORIST knowledge bases, we present a simplistic example. Consider the following *facts, hypotheses* and *observations* within the domain of educational diagnosis.

Defaults:

ASSUME visually-impaired (X)
ASSUME mentally-handicapped (X)
ASSUME arithmetic difficulties (X)
ASSUME decoding difficulties (X)

Facts:
Case-Facts:

Child's name (J), age and grade, *etc.*

Rules:

less than certain visual acuity <-> visually-impaired
less than certain score on individually-administered intelligence test <-> mentally-handicapped
poor performance <- visually-presented material and visually-impaired
poor performance <- mentally-handicapped
poor performance on arithmetic tests <- arithmetic difficulties
poor performance on reading tests <- decoding difficulties

Observations:

poor performance on an arithmetic test
poor performance on a reading test

Theories:

visually-impaired (J)
mentally-handicapped (J)
arithmetic difficulties (J) and decoding difficulties (J)

Within the domain of educational diagnosis, the defaults are possible hypotheses we are prepared to accept in a diagnosis, for example skills and subskills, misconceptions, types of errors, error patterns, ability and achievement levels. The system's facts are what is known about the domain (*rules*), as well as what is known about the particular student (*case-facts*). Case-facts include all data collected regarding the child's developmental history, family situation, previous testing, current school program, *etc.* The domain knowledge, encoded in the *rules*, includes expertise regarding educational diagnostic procedures, guidelines, legislation, terminology, as well as knowledge about the particular subjects or skills, subskills, error patterns, the interaction between errors and skills, and how such skills and patterns are measured and determined ie. knowledge regarding the interpretation of diagnostic tests.

The system builds a theory explaining a student's behaviour. The theory can be empirically verified by checking predicted behaviour against the actual performance. In our naive example above, the theory of visual-impairment makes a prediction regarding J's visual acuity: similarly, the theory regarding mental-retardation, predicts a poor performance on intelligence tests. These predictions can then be checked. In this phase, if the system does not know a fact, it can ask the user whether the fact is true. Thus the system automatically proposes experiments to test the child, so the diagnosis can be refined and so that the system can discriminate between competing diagnoses.

In the above example, one may also want to consider compound theories such as "visually-impaired (J) and arithmetic difficulties (J) and decoding difficulties (J)". There is nothing in the THEORIST philosophy prohibiting one from considering such theories. So far, we have not discussed how THEORIST might prefer one theory to another when competing theories appear to have the same predictive power. The problem of how to select the "best" theory has received some attention in the philosophy literature; for example, see (Quine and Ullian 1978) and the references therein. A common heuristic, often employed in day-to-day life, is to select the simplest (in the sense of 'most likely') or select the most modest (one theory is more modest than another if it is weaker in a logical sense). However, for certain domains, one may prefer to consider compound theories. Moreover, one's procedure for determining the "best" theory may vary, although the criteria presented in *Web of Belief* (Quine and Ullian 1978) – conservatism, modesty, simplicity, generality, refutability – can be universally applied.

3.4 How Theorist Works

THEORIST contains a goal-directed theorem prover that can determine whether a given observation logically follows from a set of axioms; the axioms consist of the facts F and the hypotheses Δ. In other words, THEORIST is trying to explain the observations G in terms of what is known (ie. the facts) and and by postulating hypotheses. Instances of the particular defaults or hypotheses used in the proof are said to constitute a candidate theory D. Therefore, from the theorem-prover the system knows that D is adequate to explain G; however, it does not yet know that $F \cup D$ is consistent. We can prove $F \cup D$ is inconsistent by using our deduction system to try to prove $F \cup D$ logically implies $\neg d$ for some $d \in D$. If this proof succeeds, the theory is not consistent. If all such proofs fail, we conclude that the theory is consistent. Trying to prove inconsistency corresponds to empirically verifying a scientific theory by allowing it to make predictions and checking the predictions against what is known.

THEORIST preserves the structure of the input supplied by the user (unlike resolution systems), and can thus explain its reasoning in the terms employed by the user. At this level it is similar to a Prolog system, but is not restricted to using Horn clauses. It can use arbitrary conjunction, disjunction, equivalence and negation.

A preliminary version of THEORIST, denoted SIMPLE THEORIST, has been implemented in Waterloo Unix Prolog (wup). SIMPLE THEORIST compiles THEORIST statements into Prolog, which is then executed like a regular wup program. This initial implementation is not intended as a final product but rather as a tool with which to experiment. Although SIMPLE THEORIST can be used to develop theories, it does not have the power of true THEORIST. It does not have a law of the excluded middle, nor does it automatically produce the contrapositive of its rules; if the contrapositives are to be used, they must be entered explicitly. Also, it does not have dependency-directed backtracking; normal Prolog backtracking is used instead. SIMPLE THEORIST does not have a theory comparitor, so it is unable to select the "best" theory; it quits after developing the first theory, unless it is subsequently directed to find additional theories. In other words, SIMPLE THEORIST can easily be forced to present all candidate theories. This restricted version of THEORIST has the advantage that it can be implemented as efficiently as Horn clauses, with the consequent gain in speed.

Unfortunately, SIMPLE THEORIST also does not include primitives for file manipulation, editing and other utilities. A mechanism exists whereby SIMPLE THEORIST can "escape" to the underlying Prolog interpreter to execute clauses that are not built into its repertoire, such as the utility primitives. The programmer may also elect to use regular Prolog for such things and use SIMPLE THEORIST for theory formulation only.

Despite the current limitations of SIMPLE THEORIST, testing of the THEORIST paradigm is underway, including the development of a small expert system for arithmetic diagnosis, which we discuss herein.

3.5 Hypothesis Abstraction Hierarchy

Although THEORIST does not require any predetermined structuring of the hypotheses, hypotheses can be arranged into an and/or tree called the ***hypothesis abstraction hierarchy*** in order to reduce the problems of combinatorial explosion. The "and-nodes" correspond to concepts that can be broken down into components. (The better theory contains more components of a composite node.) The descendents of an "or-node" are the disjoint alternatives that form more specific cases of the or-node. (The better theory contains fewer alternatives.) A node may have various facts associated with it; such facts are assertions which are relevant to the particular node. An example of an hypothesis abstraction hierarchy is discussed in Section 4.1; several other illustrations are contained in (Jones and Poole 1985).

The idea in performing a diagnosis is to move down the tree, finding more and more specific theories that are consistent and explain the observations at a greater level of detail. This has the advantage that an inappropriate subtree that could not possibly be part of a diagnosis is not considered. If there are not apparent difficulties with a particular skill, its subskills will not be investigated. In other words, an hypothesis is not considered unless it is potentially relevant.

On the other hand, if a student is experiencing difficulty with a particular skill, the corresponding node within the abstraction hierarchy will be encountered. One or more offspring of this node may in fact be considered as candidate hypotheses. This process of descending down the tree may terminate for a variety of reasons. The necessary information may not be available or all offspring of an "or-node" are possible explanations, in which case we have not gained further insight by proceeding past the "or-node". Often one can anticipate that further assessment will not yield additional information (e.g. will not distinguish among the alternative hypotheses); this type of diagnostic expertise can also be captured within THEORIST facts and hence can be exploited to prune the search.

THEORIST appears to be particularly appropriate for diagnostic expert systems as one can view the diagnosis as the expert's best theory regarding the existing problems; this is true regardless of the domain of diagnosis. Moreover the hypothesis abstraction hierarchy provides a means of modelling the case at hand at an appropriate level of abstraction. This is particularly appealing in large complex domains such as educational diagnosis. Such a structure (or structures) also allows one to break away from the too often linear organization of curricula and concentrate on task analysis. In an educational diagnosis, in order to determine what skills the student has or has not mastered, one must be able to analyze the tasks asked of the child. For any given task, one needs to specify the processes, channels and objects involved. The hypothesis abstraction hierarchies provide a framework for task analysis, that is a means of organizing the components – such as processes, channels and objects – of the tasks.

Within our expert system, several different abstraction hierarchies are employed. Most importantly, one needs to determine what processes or skills are required to accomplish this task. Consider a task which one may ask a student to perform; for example, given an arithmetic word problem, the student has to read the problem, compute and write the appropriate answer. There are several different components to this task. The main hypothesis abstraction hierarchy employed within our expert system is the ***task hierarchy*** whose root is an "and-node" with two branches *process* and *object*. For example, the arithmetic task involved in the given word problem may be that of subtraction (which is the process) of two single-digit numbers (the objects). But for any given task one may also want to specify the *mode of presentation* and the *mode of response*; hence there are two additional hierarchies that are employed for these purposes. Each such hierarchy consists of two components: channel and object. This would allow one to indicate that in the previous example, the material had been presented visually and that the required response was writing a number.

4. Experimentation with THEORIST

To evaluate the appropriateness of THEORIST, in addition to soliciting other education experts' opinions (Computer-Guided Diagnosis Workshop 1985), we have undertaken some initial experimentation. We have built a small prototype expert system for the diagnosis of arithmetic difficulties with SIMPLE THEORIST (Jones and Tubman 1986; Tubman 1986). The purpose of this experimentation was not to develop a full-blown expert system for diagnosing all aspects of arithmetic but rather to evaluate the use of THEORIST and to produce a small working prototype system.

There are several reasons for selecting arithmetic as the application domain. Although this is one area that we intend to incorporate into the final expert system, it is not our main area of concern, which is in fact reading difficulties. However, as mentioned earlier, the advantages of using a domain such as arithmetic for experimentation are numerous. In general, the processes involved in arithmetic are better understood or at least better agreed upon than those of reading or other language skills. In short there is less controversy regarding arithmetic diagnosis. Hence, it is easier to elicit the necessary diagnostic expertise and to formalize the resultant information. Also our experience has indicated that it requires less time and effort for the non-educator to understand the basics of diagnosis within the area of elementary arithmetic than that of reading. There are numerous commercially-available standardized tests for many fields of educational assessment. It is often difficult to determine precisely what such tests measure (as opposed to what they *claim* to measure) and hence it is difficult to compare information gathered from different tests. However, within the area of arithmetic, there is a great deal of agreement among the standardized tests, and also several such tests are extremely well-accepted and in common use across North America. The expert system described here uses information from two norm-referenced tests, the KeyMath Diagnostic Test and the Stanford Diagnostic Mathematics Test, and one criterion-referenced test, the Enright Diagnostic Inventory of Basic Arithmetic Skills. The latter consists of a placement test, a basic facts test, and an extremely thorough skill analysis test. Hence selecting an area such as arithmetic diagnosis for our initial experimentation has allowed us to concentrate on design and implementation issues rather than educational ones.

4.1 The Task Hierarchy

For this expert system, the task hierarchy serves as the hypothesis abstraction hierarchy. The task hierarchy is based upon a hierarchy of elementary arithmetic skills which can be characterized by the type of operation (*process*) and the type of number (*object*) involved. The task hierarchy is an and/or tree whose root is an "and-node" with two branches: *process* and *object*. However, when implementing this hierarchy, the "and-branches" have been distributed over the "or-branches".

Tasks are most general at the top of the hierarchy, and become more specialized as the levels in the tree increase. The top two levels of the hierarchy are differentiated by process; the top level task is *operation-on-numbers*, and beneath it are the tasks *addition-of-numbers*, *subtraction-of-numbers*, *multiplication-of-numbers*, and *division-of-numbers*. Directly below this level, tasks are differentiated according to the type of number; under *addition-of-numbers* there are the specializations *addition-of-integers*, *addition-of-wholes*, *addition-of-fractions* and *addition-of-decimals*. The lower levels of the hierarchy are differentiated according to such criteria as: the number of digits involved; the magnitude of the numbers, whether regrouping (borrowing and carrying) is required, and the place value of the positions used for regrouping.

Various rules and defaults are indicated at each node of the task hierarchy; this information is then exploited by THEORIST to build its theories and to control the diagnosis. Before examining how such knowledge is employed, we take a closer look at the structure of the task hierarchy.

4.2 Task Nodes

An empty node for the task hierarchy is shown in Figure 2. The node is a Prolog functor that encapsulates basic information about the task including THEORIST rules. The names of the fields and their purposes are:

name: A descriptive name of an arithmetic task.

descriptor: Contains the name of the process, a list of input objects, a list containing an output object or objects, and some explanatory text.

generalization: Holds the name field of the parent task.

specializations: Holds the names of the offspring tasks; these are more specialized versions of the current task.

queries: Contains basic background questions pertaining to the task, which are asked in the Retrospective stage of the diagnosis. Information gained from the user's responses will be asserted in THEORIST's facts database.

facts: Contains information that is known to be true for this task in any diagnosis. All facts of this type are asserted by THEORIST and stored in the facts database.

askables: Contains the names of predicates that may be proven by asking the user to supply the relevant information if it is not available from the facts or defaults. Information gained in this manner will also be asserted by THEORIST and added to the facts database.

defaults: Contains the hypotheses that can be employed to build the theories that explain the student's problem.

success: A list of names of possible theories that can be used to show a student's mastery of this task. The knowledge encoded here is, in fact, THEORIST rules which specify observations through which one can establish the specific theories. Failure to prove any success theory is *not* considered to be evidence that a skill has not been mastered.

bug: This is similar to the success field, except that the theories in its list of names are used to demonstrate that the student has not mastered a task.

prune: Stores control information in the form of a THEORIST rule; it allows the search to be cut at the appropriate point if conditions indicate that this is desireable.

prescriptions: Contains information showing what corrective measures should be undertaken to remedy a student's problems with this particular task. This information would be exploited during the Prescriptive stage.

The diagnostic expertise stored within the fields *facts*, *success*, *bug* and *prune* can all be classified as THEORIST facts, as defined in section 3.4. It is this information that the system must exploit to control the diagnostic process. The fields *facts*, *askables*, *defaults*, *success*, *bug* and *prune* may have elements that are only relevant to particular stages of the diagnosis; these are indicated accordingly.

4.3 Performing the Diagnosis

The expert system performs a diagnosis by traversing the task hierarchy four times: once for each diagnostic stage. In the Retrospective stage, the system extracts queries from the nodes of the task hierarchy, and uses them to gather background information about the child from the user. This information is asserted as SIMPLE THEORIST facts and stored in the facts database. During the Definitive stage, the system extracts facts and defaults that are applicable to this phase, and uses SIMPLE THEORIST to predict whether the student has mastered or failed to master the task. Evidence may in fact exist for both possibilities. The Analytic stage is similar to the Definitive, except that the information sought consists of more detailed explanations of why the student has not mastered the task. For example, in the Definitive phase, the system may establish that the student has difficulty with subtracting fractions; in the Analytic stage, the system may determine that this difficulty is caused by the inability to convert the fractions to a common denominator before subtracting. For our purposes of experimentation, it was not deemed necessary to implement the Prescriptive stage of the model, as a theory regarding the student's difficulties will have been developed by the end of the Analytic stage.

SIMPLE THEORIST forms its theories from facts (F) and hypotheses (Δ) to try to predict the observations (G). If a particular theory can indeed predict G, it is then considered an appropriate, although not necessarily the best, explanation. F, Δ and G are all represented by components of the task nodes in the task hierarchy.

For each task in the hierarchy, the system attempts to prove two types of theories: success theories and bug theories. If a success theory is proven, there exists evidence that the student has mastered the task. A bug theory shows that there is evidence that a task has not been mastered. These two types of theories are required because the failure to prove mastery of a task does not necessarily imply that the task has not been mastered. This can be illustrated by considering the situation where the student has not yet taken a test that measures the skill in question. No success theory can be proven, but neither can a bug theory. In this case, the information *explain-fails* is returned in place of a proven theory. Success and bug theories are defined in the *defaults* portion of the task nodes. In the domain of arithmetic diagnosis, a success theory is denoted *noproblem*, and a bug theory is denoted *problem*.

When SIMPLE THEORIST processes a node in the task hierarchy during a particular diagnostic stage, it selects each possible success theory and tries to prove it from the facts and defaults available for that task in that phase. The resulting theory, which is either a list of the defaults used in constructing the proof of the theory, or *explain-fails*, is added to the database of theories and is recorded in a file. The same procedure is then followed for the possible bug theories.

As it builds a theory, SIMPLE THEORIST attempts to use facts whenever possible. All facts are asserted in THEORIST's database of facts (F). There are, however, several different sources and types of facts. A typical fact is:

fact(is-in-grade(bryce,2) [])

This kind of fact is considered to be true throughout of the diagnosis. Facts can be pre-defined in the task hierarchy, or they can be obtained by asking the user. Facts can also be factual rules. For example, a rule that defines an elementary student as one who is in a grade from one to six inclusive is:

fact(is-elementary(Student,Grade) [ge(Grade,1) le(Grade,6)])

Another potential rule which represents the fact that criterion-referenced tests are used in the Analytic stage of diagnosis is:

fact(analytic-phase-test(Test) [criterion-referenced(Test)])

One can also assert specifics regarding a particular test or class of tests via THEORIST facts. For example, information regarding the type of scores, subtests, reliability and validity coefficients, *etc.* can all be treated as THEORIST facts. (Actually, all such data is stored in a separate database that the expert system can access.)

Facts in the real world are converted to facts in F. A diagnostician's finding that "John's stanine score in the Stanford Diagnostic Mathematics test, Red level, is 3," can also be represented as a SIMPLE THEORIST fact:

fact(score(john sdmt-red 3) [])

The defaults field contains the set of hypotheses Δ to be used to generate the candidate theories. An example of a hypothesis for a bug theory is:

Enright Wide-Range Placement Test – form A
default(problem(addition-of-wholes,Student,ewrpt-a)
[name(Student)
is-in-grade(Student,Grade)
ge(Grade,2)
le(Grade,7)
has-taken-test(Student,ewrpt-a)
error-in-ewrpt-a-questions-1-or-2])

This theory postulates that *Student* has difficulty with the task *addition-of-wholes* based on his or her results on the Enright Wide-Range Placement Test, form A. In order for this theory to be proven correct, the student must be between grades two and seven, must have taken the test, and must have had an error in questions one or two. Not all defaults are this explicit. For example, instead of specifying a particular test, an entire class of tests such as general achievement tests may be mentioned. Similarly, below average performance (e.g. one standard deviation below the mean) on the test or particular type of subtest may be the reason for believing a problem exists. Ideally, the expert system's rules are specified in terms of a test characteristics including the tasks required therein. Using this approach allows one to develop an expert system that is largely test independent. In addition to such rules, one can also include rules for specific tests if deemed necessary. Each task has many such defaults, for both the Definitive and the Analytic stages of diagnosis. If none of the theories can be proven, then *explain-fails* is returned in place of a theory.

A simple example of a more general problem that can have more than one cause is the example used earlier: difficulty with arithmetic word problems. The child can have difficulty with such problems due to his/her inability to read the problem well enough to comprehend it, or he/she may be unable to do the arithmetic task required, or both.

G. Ralph, *A Database of Diagnostic Tests and Remedial Activities*, Ministry of Education, Government of Ontario, Toronto, (1983).

Salvia, J., and Ysseldyke, J.E., *Assessment in Special and Remedial Education,* Houghton Mifflin, Boston, Mass., (1978).

Tubman, J.B., *An Expert System for Educational Diagnosis Using THEORIST*, M.Math. thesis, Department of Computer Science, University of Waterloo, (1986)

Wallace, G., and Larsen, S.C., *Educational Assessment of Learning Problems: Testing for Teaching*, Allyn and Bacon, Inc., Boston, Mass., (1979).

Young, R. M. and O'Shea, T., "Errors in Children's Subtraction", *Cognitive Science, II,* (1982), 153-177.