digital tutoring by the US Defense Advanced Research Projects agency (DARPA) has found effect sizes of 3.00 σ and 4.00 σ for one of their ITS efforts [3], but this is not the norm. Recent reviews of ITSs have found effect sizes averaging 0.70 to 0.80 σ or about the equivalent of one letter grade primarily, but not solely, in well-defined cognitive domains such as computer programming, mathematics, and physics [4], [5]. These results are similar to those produced by highly competent one-to-one human tutors.

The promise of significant growth in the effectiveness of ITSs over the next 5 to 10 years is very real. ITSs may also offer NATO substantial and unique opportunities for developing instruction that develops the critical cognitive readiness capabilities needed for military operations. Application of ITS technology in task, job, and performance aids may additionally contribute to mission success. The nature, extent, availability, and feasibility of these opportunities was identified, reviewed, and assessed for NATO applications by members of RTG HFM-237.

1.2 RTG HFM-237 OBJECTIVES

The RTG reviewed and analyzed the nature, extent, availability, and feasibility of opportunities presented by ITSs for conducting NATO education and training. While their effectiveness as one-to-one instructional tools is well documented, several barriers are challenging their widespread adoption in military training and education:

- Highly specialized skills are required to build effective ITSs – deep understanding of instructional design, one-to-one tutoring, the subject matter, and computer programming is required so the entry level to authoring ITSs is currently high.
- ITSs are expensive and time consuming to build – this limits return-on-investment for ITSs in complex domains with low density training (low throughput) low.
- The more adaptive an ITS is, the more content it needs to support tailoring and personalization of instruction – which also leads to longer development times and higher costs.
- Interoperability and data exchange between tutoring systems is severely limited by the lack of ITS standards for reuse, portability, and sharing of content, for models of learners, pedagogy, and domains, and protocols such as those for internal and external messaging.

Overcoming these challenges is essential for ITSs and their associated adaptive instructional techniques to become practical for use in large scale organizations (e.g., military schools and training centers). Each of these challenges is addressed in our recommendations in Chapter 8.

It is essential to note that ITSs are technology-based solutions for adaptive instruction in specified domains. Their practicality for use in education and training in all sectors depends on:

- Their ability to accurately model the learner and the training environment (e.g., a set of problems or immersive virtual environment).
- Their ability to effectively adapt instructional interaction to the learner's capabilities and needs.
- The likelihood that existing ITS technologies (tools and methods) may be reused and extended to support military training domains.
- The skills and cost required to author, distribute, apply, and evaluate their effectiveness.

Based on this goal of practicality, RTG-237 members canvassed adaptive training and education research and development efforts around the world to more fully understand the antecedents of effective and affordable ITSs.

sought as a way to provide every learner with an appreciable amount of the individualization needed for efficient and effective learning.

Figure 2-1 outlines a two-dimensional (content and objectives) learning space adapted from Anderson and Krathwohl [8]. It suggests a rough dichotomy of objectives with implications for how they might be learned most effectively and efficiently. Learning in most subjects begins with basic facts, nomenclature, and simple procedures that must be learned and, to an extent, memorized. Objectives for these essentials are found in the lower left-hand corner of the Anderson-Krathwohl learning space and are best presented and achieved using drill and practice as demonstrated by early research in CAI [9]-[13].

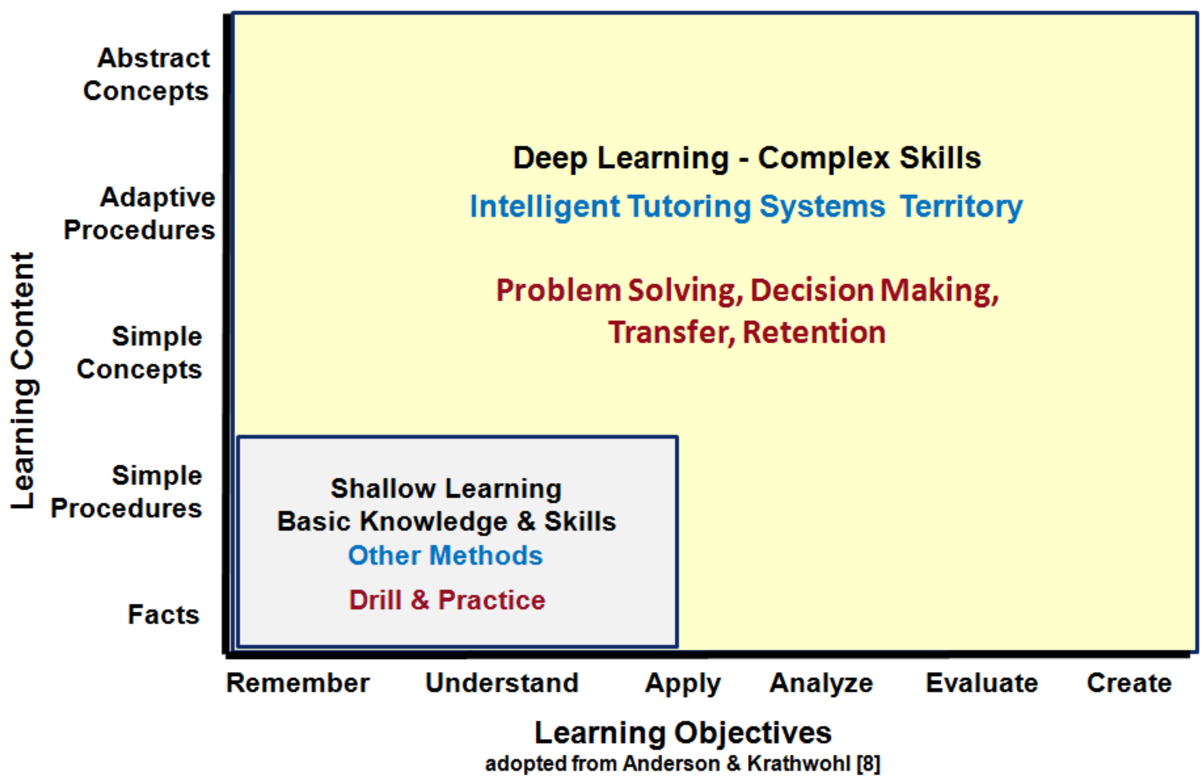

Figure 2-1: The Effective and Efficient Use of Intelligent Tutoring System Technologies.

These objectives are essential, but they do not constitute the deeper and more conceptual understanding that is required to advance from entry-level capabilities to the journeyman or expert levels needed for transfer and retention of knowledge [3], [6], [14]-[15]. Findings by these researchers and others, suggest that more relevant, persistent, and applicable material must be presented using more sophisticated techniques than drill and practice. Most of these techniques involve some form of one-on-one human tutoring, which is generally unaffordable unless it can be provided by computers that are imbued with some of the intelligence needed to clone human dialogue-based tutoring. The development of training and education based on machine or 'artificial' intelligence is, therefore, not driven by infatuation with technology or computers, but by a genuine operational need, if not an imperative, to provide a capability that is as effective and efficient as possible within the bounds and limits of resources commonly allocated to training and education.

either the learner or the tutor to initiate conversation. This approach is similar to the Socratic Method used in "How the West Was Won" [13]. GUIDON guides the learner's reasoning through a Socratic dialogue in which the tutor leads a small group in finding the answer to an open question. It uses MYCIN's 450 rules to present concrete examples. In GUIDON, as a case is selected the learner diagnosis is completed by querying the system to obtain important data and allowing the learner to propose a hypotheses.

The learner's behavior is compared to the expert behavior expected and proposed by MYCIN. The tutorial is designed to intervene when the learner asks for help or when their actions are incorrect (e.g., buggy tutoring). For example, GUIDON might suggest to the learner the next steps they should take or advise them that their questions are not relevant to the case being presented. GUIDON had the capability to understand simple sentences and a list of standard commands (e.g., "help", "justify", "summarize"). The feedback to the learner was generated from a store of primitive phrases so in a basic sense GUIDON it was generative; it could produce responses. The learner is allowed to take the initiative to change the topic. This requires the program to respond intelligently by maintaining a record of the context of the interaction (e.g. conditions).

Although GUIDON was viewed as one of the most sophisticated ITSs ever built, it was not as effective as expected. This was largely due to the way its expertise was represented. While the knowledge was accurate and at the right level of granularity, the schema did not reflect the way that expert human diagnosticians actually reasoned so the system's communication incompatible with medical diagnosis practices of the day. Clancey [14]-[16] reconfigured MYCIN to NEOMYCIN, where GUIDON expressed tutoring strategies in terms of an abstract representational scheme. NEOMYCIN separated the reasoning strategy and expressed abstractly tutoring strategies, in terms of tasks that manipulate specific types of information. Clancey built NEOMYCIN with the ability to justify the reasoning processes and interact with users in the terms they could relate to and understand. One of the goals for NEOMYCIN was to develop a generic tutoring paradigm for all of the classes of problem solvers. He defined these classes of problem solvers as heuristic classification. He classified a given case according to a predetermined taxonomy by relating some features of the data to descriptions of the candidate categories. The intention was to make sure that such a system was based on a thorough understanding of the problem–solving process. Clancey also went on to develop HERACLES a generic system for a class of problem-solving strategies that were used to define and create knowledge engineering tools.

In summary, GUIDON was a unique attempt to turn an existing expert system for a complex task into a medical diagnosis tutor. A major contribution was the identification and separate treatment of different types of knowledge that must be made available in order for a tutor to function effectively. Second, the development of GUIDON identified the need to develop an expert system that contained pedagogical expertise. Clancey called this the *tutorial module*. Ideally, the tutorial strategies developed and delivered by the tutorial module should be domain-independent in that they are used to guide learning in a variety of domains. Although, the implementation of the rule-base tutorial module in NEOMYCIN was deemed effective, the use of MYCIN as a source of domain expertise proved to be problematic as its reasoning strategy was inflexible and it had difficulty tracking the learner's progress. One of the lessons-learned from this research revealed some of the limitations of using compiled production systems for instructional purposes. However, the overall result of this project was a new model of diagnostic thinking and it provided the foundation for many of the medical and non-medical (e.g., Science, Technology, Engineering, and Mathematics − STEM) tutors today.

7.2.1.2 Methods for Drawing Conclusions in Diagnostic Reasoning

The method described here is important to ITS development in that it describes an early study where standardized errors were identified and used to judge the performance of the learner as an expert model. The study by Voytovich and Suffredini [2] explored the characteristics of premature diagnostic conclusions