AGI Preschool:
A Framework for Evaluating Early-Stage Human-like AGIs

Ben Goertzel, Stephan Vladimir Bugaj

Novamente LLC & Singularity Institute for AI, AGI Research Institute
1405 Bernerd Place, Rockville MD 20851, USA
ben@goertzel.org, stephan@bugaj.com

Abstract

A class of environments for teaching and evaluating AGI systems is described, modeled loosely on preschools used for teaching human children and intended specifically for early-stage systems aimed at approaching human-level, human-inspired AGI through a process resembling human developmental psychology. Some particulars regarding implementation of environments in this class in online virtual worlds are presented, along with details on associated evaluation tasks, and discussion of environments potentially appropriate for "AGI preschool graduates," based on more advanced human schooling.

Introduction

One of the many difficult issues arising in the course of research on human-level AGI is that of "evaluation and metrics" – i.e., AGI intelligence testing. The Turing test (Turing, 1950) approaches the question of how to tell if an AGI has achieved human-level intelligence, and while there are debates and subtleties even with this well-known test (Moor, 200l; French, 1990; Hayes and Ford, 1995), the question we'll address here is a significantly trickier one: assessing the quality of *incremental* progress toward human-level AI.

 (Laird, et al, 2009) discusses some of the general difficulties involved in this type of assessment, and some requirements that any viable approach must fulfill. Here, rather than surveying the spectrum of possibilities, we will focus on describing in detail one promising approach: emulation, in a multiuser online virtual world, of an environment similar to preschools used in early human childhood education. Complete specification of an "AGI Preschool" would require much more than a brief conference paper; our goal here is to sketch the idea in broad outline, and give a few examples of the types of opportunities such an environment would afford for instruction, spontaneous learning and formal and informal evaluation of certain sorts of early-stage AGI systems.

The Need for "Realistic" AGI Testing

One might question the need or importance of a new, overall framework for AGI intelligence testing, such as the AGI Preschool appraoch described here. After all, there has already been a lot of work on evaluating the capability of various AI systems and techniques. However, we believe that the approaches typically taken have significant shortcomings from an AGI perspective.

Certainly, the AI field has inspired many competitions, each of which tests some particular type or aspect of intelligent behavior. Examples include robot competitions, tournaments of computer chess, poker, backgammon and so forth at computer olympiads, trading-agent competition, language and reasoning competitions like the Pascal Textual Entailment Challenge, and so on. In addition to these, there are many standard domains and problems used in the AI literature that are meant to capture the essential difficulties in a certain class of learning problems: standard datasets for face recognition, text parsing, supervised classification, theorem-proving, question-answering and so forth.

However, the value of these sorts of tests for AGI is predicated on the hypothesis that the degree of success of an AI program at carrying out some domain-specific task, is correlated the the potential of that program for being developed into a robust AGI program with broad intelligence. If AGI and problem-area-specific "narrow AI" are in fact very different sorts of pursuits requiring very different principles, as we suspect (Goertzel and Pennachin, 2006), then these tests are not strongly relevant to the AGI problem.

There are also some standard evaluation paradigms aimed at AI going beyond specific tasks. For instance, there is a literature on "multitask learning," where the goal for an AI is to learn one task quicker given another task solved previously (Caruna, 1997; Thrun and Mitchell, 1995; Ben-David and Schuller, 2003; Taylor and Stone, 2007). This is one of the capabilities an AI agent will need to simultaneously learn different types of tasks as proposed in the Preschool scenario given here. And there is a literature on "shaping," where the idea is to build up the capability of an AI by training it on progressively more difficult versions of the same tasks (Laud and Dejong, 203; Walsh and Littman, 2006). Again, this is one sort of capability an AI will need to possess if it is to move up some type of curriculum, such as a school curriculum.

While we applaud the work done on multitask learning

and shaping, we feel that exploring these processes using mathematical abstractions, or in the domain of various machine-learning or robotics test problems, may not adequately address the problem of AGI. The problem is that generalization among tasks, or from simpler to more difficult versions of the same task, is a process whose nature may depend strongly on the overall nature of the set of tasks and task-versions involved. Real-world tasks have a subtlety of interconnectedness and developmental course that is not captured in current mathematical learning frameworks nor standard AI test problems.

To put it mathematically, we suggest that the universe of real-world human tasks has a host of "special statistical properties" that have implications regarding what sorts of AI programs will be most suitable; and that, while exploring and formalizing the nature of these statistical properties is important, an easier and more reliable approach to AGI testing is to create a testing environment that embodies these properties implicitly, via its being an emulation of the cognitively meaningful aspects of the real-world human learning environment.

One way to see this point vividly is to contrast the current proposal with the "General Game Player" AI competition, in which AIs seek to learn to play games based on formal descriptions of the rules.[1] Clearly doing GGP well requires powerful AGI; and doing GGP even mediocrely probably requires robust multitask learning and shaping. But we suspect GGP is far inferior to AGI Preschool as an approach to testing early-stage AI programs aimed at roughly humanlike intelligence. This is because, unlike the tasks involved in AI Preschool, the tasks involved in doing simple instances of GGP seem to have little relationship to humanlike intelligence or real-world human tasks.

Multiple Intelligences

Intelligence testing is, we suggest, best discussed and pursued in the context of a theoretical interpretation of "intelligence." As there is yet no consensus on the best such interpretation (Legg and Hutter (2006) present a summary of over 70 definitions of intelligence presented in the research literature), this is a somewhat subtle point.

In our own prior work we have articulated a theory based on the notion of intelligence as "the ability to achieve complex goals in complex environments," a definition that relates closely to Legg and Hutter's (2007) more rigorous definition in terms of statistical decision theory. However, applying this sort of theory to practical intelligence testing seems very difficult, in that it requires an assessment of the comparative complexity of various real-world tasks and environments. As real-world tasks and environments are rarely well-formalized, one's only pragmatic option is to assess complexity in terms of human common sense or natural language, which is an approach fraught with "hidden rocks," though it might prove fruitful

if pursued with sufficient care and effort.

Here we have chosen an alternate, complementary approach, choosing as our inspiration Gardner's (1983) multiple intelligences (MI) framework -- a psychological approach to intelligence assessment based on the idea that different people have mental strengths in different high-level domains, so that intelligence tests should contain aspects that focus on each of these domains separately. MI does not contradict the "complex goals in complex environments" view of intelligence, but rather may be interpreted as making specific commitments regarding which complex tasks and which complex environments are most important for roughly human-like intelligence.

MI does not seek an extreme generality, in the sense that it explicitly focuses on domains in which humans have strong innate capability as well as general-intelligence capability; there could easily be non-human intelligences that would exceed humans according to both the commonsense human notion of "general intelligence" and the generic "complex goals in complex environments" or Hutter/Legg-style definitions, yet would not equal humans on the MI criteria. This strong anthropocentrism of MI is not a problem from an AGI perspective so long as one uses MI in an appropriate way, i.e. only for assessing the extent to which an AGI system displays specifically *human-like* general intelligence. This restrictiveness is the price one pays for having an easily articulable and relatively easily implementable evaluation framework.

Table 1 sumarizes the types of intelligence included in Gardner's MI theory. Later on, we will suggest that one way to assess an AGI Preschool implementation is to ask how well it covers all the bases outlined in MI theory.

Intelligence Type	Aspects
Linguistic	words and language, written and spoken; retention, interpretation and explanation of ideas and information via language, understands relationship between communication and meaning
Logical-Mathematical	logical thinking, detecting patterns, scientific reasoning and deduction; analyse problems, perform mathematical calculations, understands relationship between cause and effect towards a tangible outcome
Musical	musical ability, awareness, appreciation and use of sound; recognition of tonal and rhythmic patterns, understands relationship between sound and feeling
Bodily-Kinesthetic	body movement control, manual dexterity, physical agility and balance; eye and body coordination
Spatial-Visual	visual and spatial perception; interpretation and creation of images; pictorial imagination and expression; understands relationship between images and meanings, and between space and effect
Interpersonal	perception of other people's feelings; relates to others; interpretation of behaviour and communications; understands relationships between people and their situations

Table 1. Types of Intelligence, Aspects and Testing

---

[1] http://games.stanford.edu/

Elements of Preschool Design

What we mean by an "AGI Preschool" is simply a porting to the AGI domain of the essential aspects of human preschools. While there is significant variance among preschools there are also strong commonalities, grounded in educational theory and experience. We will briefly discuss both the physical design and educational curriculum of the typical human preschool, and which aspects transfer effectively to the AGI context.

On the physical side, the key notion in modern preschool design is the "learning center," an area designed and outfitted with appropriate materials for teaching a specific skill. Learning centers are designed to encourage learning by doing, which greatly facilitates learning processes based on reinforcement, imitation and correction (see Goertzel et al (2008) for a discussion of the importance of this combination in an AGI context); and also to provide multiple techniques for teaching the same skills, to accommodate different learning styles and prevent over-fitting and overspecialization in the learning of new skills.

Centers are also designed to cross-develop related skills. A "manipulatives center," for example, provides physical objects such as drawing implements, toys and puzzles, to facilitate development of motor manipulation, visual discrimination, and (through sequencing and classification games) basic logical reasoning. A "dramatics center," on the other hand, cross-trains interpersonal and empathetic skills along with bodily-kinesthetic, linguistic, and musical skills. Other centers, such as art, reading, writing, science and math centers are also designed to train not just one area, but to center around a primary intelligence type while also cross-developing related areas. For specific examples of the learning centers associated with particular contemporary preschools, see (Neilsen, 2006).

In many progressive, student-centered preschools, students are left largely to their own devices to move from one center to another throughout the preschool room. Generally, each center will be staffed by an instructor at some points in the day but not others, providing a variety of learning experiences. At some preschools students will be strongly encouraged to distribute their time relatively evenly among the different learning centers, or to focus on those learning centers corresponding to their particular strengths and/or weaknesses.

Elements of Preschool Curriculum

While preschool curricula vary considerably based on educational philosophy and regional and cultural factors, there is a great deal of common, shared wisdom regarding the most useful topics and methods for preschool teaching. Guided experiential learning in diverse environments and using varied materials is generally agreed upon as being an optimal methodology to reach a wide variety of learning types and capabilities. Hands-on learning provides grounding in specifics, where as a diversity of approaches allows for generalization.

Core knowledge domains are also relatively consistent, even across various philosophies and regions. Language, movement and coordination, autonomous judgment, social skills, work habits, temporal orientation, spatial orientation, mathematics, science, music, visual arts, and dramatics are universal areas of learning which all early childhood learning touches upon. The particulars of these skills may vary, but all human children are taught to function in these domains. The level of competency developed may vary, but general domain knowledge is provided. For example, most kids won't be the next Maria Callas, Ravi Shankar or Gene Ween, but nearly all learn to hear, understand and appreciate music.

Type of Capability	Specific Skills to be Evaluated
Story Understanding	• Understanding narrative sequence • Understanding character development • Dramatize a story • Predict what comes next in a story
Linguistic	• Give simple descriptions of events • Describe similarities and differences • Describe objects and their functions
Linguistic / Spatial-Visual	• Interpreting pictures
Linguistic / Social	• Asking questions appropriately • Answering questions appropriately • Talk about own discoveries • Initiate conversations • Settle disagreements • Verbally express empathy • Ask for help • Follow directions
Linguistic / Scientific	• Provide possible explanations for events or phenomena • Carefully describe observations • Draw conclusions from observations
Logical-Mathematical	• Categorizing • Sorting • Arithmetic • Performing simple "proto-scientific experiments"
Nonverbal Communication	• Communicating via gesture • Dramatizing situations • Dramatizing needs, wants • Express empathy
Spatial-Visual	• Visual patterning • Self-expression through drawing • Navigate
Objective	• Assembling objects • Disassembling objects • Measurement • Symmetry • Similarity between structures (e.g. block structures and real ones)
Interpersonal	• Cooperation • Display appropriate behavior in various settings • Clean up belongings • Share supplies
Emotional	• Delay gratification • Control emotional reactions • Complete projects

Table 2. Preschool cognitive tests

Table 2 reviews the key capabilities taught in preschools, and identifies the most important specific skills that need to be evaluated in the context of each capability. This table was assembled via surveying the curricula from a number of currently existing preschools employing different methodologies both based on formal academic cognitive theories (Schunk 2007) and more pragmatic approaches, such as: Montessori (Montessori, 1912), Waldorf (Steiner 2003), Brain Gym (www.braingym.org) and Core Knowledge (www.coreknowledge.org).

Preschool in the Light of Intelligence Theory

Comparing Table 2 to the Multiple Intelligences framework, the high degree of harmony is obvious, and is borne out by more detailed analysis which is omitted here for space reasons. Preschool curriculum as standardly practiced is very well attuned to MI, and naturally covers all the bases that Gardner identifies as important. And this is not at all surprising since one of Gardner's key motivations in articulating MI theory was the pragmatics of educating humans with diverse strengths and weaknesses.

Regarding intelligence as "the ability to achieve complex goals in complex environments," it is apparent that preschools are specifically designed to pack a large variety of different micro-environments (the learning centers) into a single room, and to present a variety of different tasks in each environment. The environments constituted by preschool learning centers are designed as microcosms of the most important aspects of the environments faced by humans in their everyday lives.

Task-Based Assessment in AGI Preschool

Professional pedagogues such as (Chen & McNamee, 2007) discuss evaluation of early childhood learning as intended to assess both specific curriculum content knowledge as well as the child's learning process. It should be as unobtrusive as possible, so that it just seems like another engaging activity, and the results used to tailor the teaching regimen to use different techniques to address weaknesses and reinforce strengths.

For example, with group building of a model car, students are tested on a variety of skills: procedural understanding, visual acuity, motor acuity, creative problem solving, interpersonal communications, empathy, patience, manners, and so on. With this kind of complex, yet engaging, activity as a metric the teacher can see how each student approaches the process of understanding each subtask, and subsequently guide each student's focus differently depending on strengths and weaknesses.

Next we describe some particular tasks that AGIs may be meaningfully assigned in the context of a general AGI Preschool design and curriculum as described above. Due to length limitations this is a very partial list, and is intended as evocative rather than comprehensive.

Any one of these tasks can be turned into a rigorous quantitative test, thus allowing the precise comparison of different AGI systems' capabilities; but we have chosen not to emphasize this point here, partly for space reasons and partly for philosophical ones. In some contexts the quantitative comparison of different systems may be the right thing to do, but as discussed in (Laird et al, 2008) there are also risks associated with this approach, including the emergence of an overly metrics-focused "bakeoff mentality" among system developers, and overfitting of AI abilities to test taking. What is most important is the isolation of specific tasks on which different systems may be experientially trained and then qualitatively assessed and compared, rather than the evaluation of quantitative metrics.

Table 3 lists a sampling of different tests for each intelligence type to be assessed.

Intelligence Type	Test
Linguistic	write a set of instructions; speak on a subject; edit a written piece or work; write a speech; commentate on an event; apply positive or negative 'spin' to a story
Logical-Mathematical	perform arithmetic calculations; create a process to measure something; analyse how a machine works; create a process; devise a strategy to achieve an aim; assess the value of a proposition
Musical	perform a musical piece; sing a song; review a musical work; coach someone to play a musical instrument
Bodily-Kinesthetic	juggle; demonstrate a sports technique; flip a beer-mat; create a mime to explain something; toss a pancake; fly a kite
Spatial-Visual	design a costume; interpret a painting; create a room layout; create a corporate logo; design a building; pack a suitcase or the boot of a car
Interpersonal	interpret moods from facial expressions; demonstrate feelings through body language; affect the feelings of others in a planned way; coach or counsel another

Table 3. Prototypical intelligence assessment tasks.

Task-oriented testing allows for feedback on applications of general pedagogical principles to real-world, embodied activities. This allows for iterative refinement based learning (shaping), and cross development of knowledge acquisition and application (multitask learning). It also helps militate against both cheating, and over-fitting, as teachers can make ad-hoc modifications to the tests to determine if this is happening and correct for it if necessary.

E.g., consider a linguistic task in which the AGI is required to formulate a set of instructions encapsulating a given behavior (which may include components that are physical, social, linguistic, etc.). Note that although this is presented as centrally a linguistic task, it actually involves a diverse set of competencies since the behavior to be described may encompass multiple real-world aspects.

To turn this task into a more thorough test one might involve a number of human teachers and a number of human students. Before the test, an ensemble of copies of the AGI would be created, with identical knowledge state. Each copy would interact with a different human teacher, who would demonstrate to it a certain behavior. After testing the AGI on its own knowledge of the material, the teacher would then inform the AGI that it will then be tested on its ability to verbally describe this behavior to another. Then, the teacher goes away and the copy interacts with a series of students, attempting to convey to the students the instructions given by the teacher.

The teacher can thereby assess both the AGI's understanding of the material, and the ability to explain it to the other students. This separates out assessment of understanding from assessment of ability to communicate understanding, attempting to avoid conflation of one with the other. The design of the training and testing needs to account for potential

This testing protocol abstracts away from the particularities of any one teacher or student, and focuses on effectiveness of communication in a human context rather than according to formalized criteria. This is very much in the spirit of how assessment takes place in human preschools (with the exception of the copying aspect): formal exams are rarely given in preschool, but pragmatic, socially-embedded assessments are regularly made.

By including the copying aspect, more rigorous statistical assessments can be made regarding efficacy of different approaches for a given AGI design, independent of past teaching experiences. The multiple copies may, depending on the AGI system design, then be able to be reintegrated, and further "learning" be done by higher-order cognitive systems in the AGI that integrate the disparate experiences of the multiple copies.

This kind of parallel learning is different from both sequential learning that humans do, and parallel presences of a single copy of an AGI (such as in multiple chat rooms type experiments). All three approaches are worthy of study, to determine under what circumstances, and with which AGI designs, one is more successful than another.

It is also worth observing how this test could be tweaked to yield a test of generalization ability. After passing the above, the AGI could then be given a description of a new task (acquisition), and asked to explain the new one (variation). And, part of the training behavior might be carried out unobserved by the AGI, thus requiring the AGI to infer the omitted parts of the task it needs to describe.

Another popular form of early childhood testing is puzzle block games. These kinds of games can be used to assess a variety of important cognitive skills, and to do so in a fun way that not only examines but also encourages creativity and flexible thinking. Types of games include pattern matching games in which students replicate patterns described visually or verbally, pattern creation games in which students create new patterns guided by visually or verbally described principles, creative interpretation of patterns in which students find meaning in the forms, and free-form creation. Such games may be individual or cooperative.

Cross training and assessment of a variety of skills occurs with pattern block games: for example, interpretation of visual or linguistic instructions, logical procedure and pattern following, categorizing, sorting, general problem solving, creative interpretation, experimentation, and kinematic acuity. By making the games cooperative, various interpersonal skills involving communication and cooperation are also added to the mix.

The puzzle block context bring up some general observations about the role of kinematic and visuospatial intelligence in the AGI Preschool. Outside of robotics and computer vision, AI research has often downplayed these sorts of intelligence (though, admittedly, this is changing in recent years, e.g. with increasing research focus on diagrammatic reasoning). But these abilities are not only necessary to navigate real (or virtual) spatial environments. They are also important components of a coherent, conceptually well-formed understanding of the world in which the student is embodied. Integrative training and assessment of both rigorous cognitive abilities generally most associated with both AI and "proper schooling" (such as linguistic and logical skills) along with kinematic and aesthetic/sensory abilities is essential to the development of an intelligence that can successfully both operate in and sensibly communicate about the real world in a roughly humanlike manner. Whether or not an AGI is targeted to interpret physical-world spatial data and perform tasks via robotics, in order to communicate ideas about a vast array of topics of interest to any intelligence in this world, an AGI must develop aspects of intelligence other than logical and linguistic cognition.

Issues with Virtual Preschool Engineering

One technical point that has come to our attention in exploring the AGI Preschool concept pertains to the available infrastructure for creating such a framework. Current standard methods of controlling avatar behaviors and avatar-object interactions in existing virtual environment engines are not sufficiently flexible to fulfill the requirements of an AGI Preschool. Such flexibility is feasible using current technology, but has not been implemented due to lack of motivation.

The most glaring deficit in current platforms is the lack of flexibility in terms of tool use. In most of these systems today, an avatar can pick up or utilize an object, or two objects can interact, only in specific, pre-programmed ways. For instance, an avatar might be able to pick up a virtual screwdriver only by the handle, rather than by pinching the blade between its fingers. This places severe limits on creative use of tools, which is absolutely critical in a preschool context. The solution to this problem is clear: adapt existing generalized physics engines to mediate avatar-object and object-object interactions. This would require more computation than current approaches, but not more than is feasible in a research context.

One way to achieve this goal would be to integrate a robot simulator with a virtual world or game engine, for instance to modify the OpenSim (opensimulator.org) virtual world to use the Gazebo (playerstage.sourceforge.net) robot simulator in place of its current physics engine. While tractable, such a project would require considerable software engineering effort.

Beyond Preschool

Once an AGI passes preschool, what are the next steps? There is still a long way to go, from preschool to an AGI system that is capable of, say, passing the Turing Test or serving as an effective artificial scientist.

Our suggestion is to extend the school metaphor further, and make use of existing curricula for higher levels of virtual education: grade school, secondary school, and all levels of post-secondary education. If an AGI can pass online primary and secondary schools such as e-tutor.com, and go on to earn an online degree from an accredited university, then clearly said AGI has successfully achieved "human level, roughly humanlike AGI." This sort of testing is interesting not only because it allows assessment of stages intermediate between preschool and adult, but also because it tests humanlike intelligence without requiring precise imitation of human behavior.

If an AI can get a BA degree at an accredited university, via online coursework (assuming for simplicity courses where no voice interaction is needed), then we should consider that AI to have human-level intelligence. University coursework spans multiple disciplines, and the details of the homework assignments and exams are not known in advance, so like a human student the AGI team can't cheat.

In addition to the core coursework, a schooling approach also tests basic social interaction and natural language communication, ability to do online research, and general problem solving ability. However, there is no rigid requirement to be strictly humanlike in order to pass university classes.

References

Ben-David, S., Schuller, R.: Exploiting task relatedness for learning multiple tasks. In: Proceedings of the 16th Annual Conference on Learning Theory. (2003)

Caruana, R. (1997). Multitask learning. Machine Learning, 28, 41–75.

Chen, Jie-Qi and Gillian McNamee (2007). Bridging: Assessment for Teaching and Learning in Early Childhood Classrooms., Corwin Press.

French, R. (1990) "Subcognition and the Limits of the Turing Test" Mind 99, 53-65.Gardner, Howard. (1983) "Frames of Mind: The Theory of Multiple Intelligences." New York: Basic Books.

Goertzel, Ben, Cassio Pennachin, Nil Geissweiller, Moshe Looks, Andre Senna, Welter Silva, Ari Heljakka, Carlos Lopes (2008). An Integrative Methodology for Teaching Embodied Non-Linguistic Agents, Applied to Virtual Animals in Second Life. Proceedings of the First Conference on Artificial General Intelligence, IOS Press.

Goertzel, Ben and Stephan Vladimir Bugaj (2006). Stages of Cognitive Development in Uncertain-Logic-Based AI Systems, in Proceedings of First Conference on Artificial General Intelligence, IOS Press

Goertzel, Ben, Ari Heljakka, Stephan Vladimir Bugaj, Cassio Pennachin, and Moshe Looks (2006). Exploring Android Developmental Psychology in a Simulation World. Presented at ICCS/CogSci 2006 Android Science Workshop.

Goertzel, Ben, Ken Silverman, Cate Harteley, Stephan Vladimir Bugaj and Mike Ross. The Baby Webmind Project. Presented at AISB 2000.

Hayes, P., and Ford, K. (1995)"Turing Test Considered Harmful" Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence Montreal, 972-7.

Laird, John, Robert Wray, Robert Marinier and Pat Langley (2009). Claims and Challenges in Evaluating Human-Level Intelligent Systems, Proceedings of AGI-09

Laud, A., & Dejong, G. (2003). The influence of reward on the speed of reinforcement learning: An analysis of shaping. Proceedings of the 20th international conference on machine learning.

Legg, Shane and Marcus Hutter (2007). Universal Intelligence: A Definition of Machine Intelligence. In Minds and Machines, pages 391-444, volume 17, number 4, November 2007.

Legg, Shane and Marcus Hutter (2006). A Collection of Definitions of Intelligence In B. Goertzel, editor, Advances in Artificial General Intelligence, IOS Press, 2006.

Li, L., Walsh, T., & Littman, M. (2006). Towards a unified theory of state abstraction for MDPs. Proceedings of the ninth international symposium on AI and mathematics.

Moor, J. (2001) "The Status and Future of the Turing Test" Minds and Machines 11, 77-93.

Montessori, Maria (1912). The Montessori Method, Frederick A. Stokes Company,

Neilsen, Dianne Miller (2006). Teaching Young Children, Preschool-K: A Guide to Planning Your Curriculum, Teaching Through Learning Centers, and Just About Everything Else, Corwin Press.

Reece, Charlotte Strange, Kamii, Constance (2001). The measurement of volume: Why do young children measure inaccurately?, School Science and Mathematics, Nov 2001,

Schunk, Dale (2007). Theories of Learning: An Educational Perspective, Prentice Hall.

Steiner, Rudolf and S.K. Sagarin. (2003) What is Waldorf Education? Three Lectures, Steiner Books.

Taylor, M., & Stone, P. (2007). Cross-domain transfer for reinforcement learning. Proceedings of the 24th International Conference on Machine Learning.

Thrun, S., & Mitchell, T. (1995). Lifelong robot learning. Robotics and Autonomous Systems, 15, 25–46.

Turing, Alan (October 1950), "Computing Machinery and Intelligence", Mind LIX(236): 433–460,