

# My Computer Is an Honor Student — but How Intelligent Is It? Standardized Tests as a Measure of AI

**Peter Clark**

Allen Institute for Artificial Intelligence

**Oren Etzioni**

Allen Institute for Artificial Intelligence

DOI: <https://doi.org/10.1609/aimag.v37i1.2636>

## Abstract

Given the well-known limitations of the Turing Test, there is a need for objective tests to both focus attention on, and measure progress towards, the goals of AI. In this paper we argue that machine performance on standardized tests should be a key component of any new measure of AI, because attaining a high level of performance requires solving significant AI problems involving language understanding and world modeling - critical skills for any machine that lays claim to intelligence. In addition, standardized tests have all the basic requirements of a practical test: they are accessible, easily comprehensible, clearly measurable, and offer a graduated progression from simple tasks to those requiring deep understanding of the world. Here we propose this task as a challenge problem for the community, summarize our state-of-the-art results on math and science tests, and provide supporting datasets



Published

2016-04-13

How to Cite

Clark, P., & Etzioni, O. (2016). My Computer Is an Honor Student — but How Intelligent Is It? Standardized Tests as a Measure of AI. *AI Magazine*, 37(1), 5-12.

<https://doi.org/10.1609/aimag.v37i1.2636>

More Citation Formats

Issue

[Vol. 37 No. 1: Spring 2016](#)

Section

Articles

License

Authors who publish with this journal agree to the following terms:

1. Author(s) agree to transfer their copyrights in their article/paper to the Association for the Advancement of Artificial Intelligence (AAAI), in order to deal with future requests for reprints, translations, anthologies, reproductions, excerpts, and other publications. This grant will include, without limitation, the entire copyright in the article/paper in all countries of the world, including all renewals, extensions, and reversions thereof, whether such rights current exist or hereafter come into effect, and also the exclusive right to create electronic versions of the article/paper, to the extent that such right is not subsumed under copyright.
2. The author(s) warrants that they are the sole author and owner of the copyright in the above article/paper, except for those portions shown to be in quotations; that the article/paper is original throughout; and that the undersigned right to make the grants set forth above is complete and unencumbered.
3. The author(s) agree that if anyone brings any claim or action alleging facts that, if true, constitute a breach of any of the foregoing warranties, the author(s) will hold harmless and indemnify AAAI, their grantees, their licensees, and their distributors against any liability, whether under judgment, decree, or compromise, and any legal fees and expenses arising out of that claim or actions, and the undersigned will cooperate fully in any defense AAAI may make to such claim or action. Moreover, the undersigned agrees to cooperate in any claim or other action seeking to protect or enforce any right the undersigned has granted to AAAI in the article/paper. If any such claim or action fails because of facts that constitute a breach of any of the foregoing warranties, the undersigned agrees to reimburse whomever brings such claim or action for expenses and attorneys' fees incurred therein.
4. Author(s) retain all proprietary rights other than copyright (such as patent rights).
5. Author(s) may make personal reuse of all or portions of the above article/paper in other works of their own authorship.
6. Author(s) may reproduce, or have reproduced, their article/paper for the author's personal use, or for company use provided that AAAI copyright and the source are indicated, and that the copies are not used in a way that implies AAAI endorsement of a product or service of an employer, and that the copies per se are not offered for sale. The foregoing right shall not permit the posting of the article/paper in electronic or digital form on any computer network, except by the author or the author's employer, and then only on the author's or the employer's own web page or ftp site. Such web page or ftp site, in addition to the aforementioned requirements of this Paragraph, must provide an electronic reference or link back to the AAAI electronic server, and shall not post other AAAI copyrighted materials not of the author's or the employer's creation (including tables of contents with links to other papers) without AAAI's written permission.
7. Author(s) may make limited distribution of all or portions of their article/paper prior to publication.
8. In the case of work performed under U.S. Government contract, AAAI grants the U.S. Government royalty-free permission to reproduce all or portions of the above article/paper, and to authorize others to do so, for U.S. Government purposes.
9. In the event the above article/paper is not accepted and published by AAAI, or is withdrawn by the author(s) before acceptance by AAAI, this agreement becomes null and void.

## Information

[For Readers](#)

[For Authors](#)

[Open Journal Systems](#)

Part of the

[PKP Publishing Services Network](#)

Copyright © 2019, Association for the Advancement of Artificial Intelligence. All rights reserved.

Platform &  
workflow by  
**OJS / PKP**

Standardized Tests as a Measure of AI Peter Clark and Oren Etzioni Abstract Given the well-known limitations of the Turing Test, there is a need for objective tests to both focus attention on, and measure progress towards, the goals of AI. This continuum is highly desirable, as it means that there is a low barrier to entry, allowing researchers to make initial inroads into the task, while significant AI challenges need to be solved to do well in the exam. The diversity of questions also ensures a variety of skills are tested for, and guards against finding a simple shortcut that may answer them all without requiring any depth of understanding. The student feels the objects but does not look into the bag. Which property of the objects can the student most likely identify? Standardized Tests as a Measure of AI. By Clark, Peter; Etzioni, Oren. Read preview. Magazine article AI Magazine. In this article, we argue that standardized tests are an effective and practical assessment of many aspects of machine intelligence, and should be part of any

comprehensive measure of AI progress. While a crisp definition of machine intelligence remains elusive, we can enumerate some general properties we might expect of an intelligent machine. The list is potentially long (for example, Legg and Hutter [2007]), but should at least include the ability to (1) answer a wide variety of questions, (2) answer complex questions, (3) demonstrate commonsense and world knowledge, and (4) acquire new knowledge. Intelligent and intelligence are not quite the same thing. One relates to application, the other to potential. That is why possessing a high IQ, although it would imply that they are intelligent, really implies potential. Unless they apply that IQ then it just remains a potential. However, these tests do not predict how an individual will utilize the potential with which they were born. Some may maximize it, others may squander it. For instance, we can measure how big a person's short-term memory capacity is by asking them to read strings of digits or words and then repeat them from memory. Still the results of that test could vary based on the physical condition of an individual at the moment of testing (cold, hot, hungry, etc.). Most standardized tests are created to evaluate the information individual students have retained in core competency areas of learning. Although many students tend to find classes like art, music, and physical education to be the most enjoyable at school, there is also a need for reading, writing, and arithmetic.

2. It can be very expensive for some states. Standardized testing is an assessment tool that is reasonably affordable for most geographic locations. New York State, for example, spends just \$7 per student on standardized tests. Oregon spends \$13 per student, while Georgia spends \$14 per student. Measuring them through standardized testing allows teachers, parents, and administrators to understand where each child is at on their own learning path.