High Stakes Tests Don't Belong in Science Museums: We Can Do Better Than That!

George E. Hein
Professor Emeritus, Lesley University

Presented at a panel, "Can Informal Science and Mathematics Learning Coexist with High-Stakes Testing?" at ASTC annual Meeting, Phoenix, AZ, October 10, 2001

A. HISTORY

Large scale, standardized, paper-and-pencil testing, especially for "high-stakes"—such as public school graduation, assignment to special education classes, college admission, draft/military status or job placement—has a tragic, anti-democratic and shoddy history.[1]

Tragic because it has repeatedly been used to defend racist policies; anti-democratic because it has been cited for almost a century to support discriminatory policies in education and job placement; and shoddy because it illustrates the defense of strong, definitive but unwarranted conclusions based on poor science and mathematics. It's failures, on both political and intellectual grounds, include unsubstantiated claims concerning:

- Inherent intelligence of different ethnic groups
- Educability (or lack of it) of individual students
- Immutability of test scores
- Generalizability of results to real-life situations.

Thus, any use of such tests to determine life opportunities needs to be examined critically.[2]

B. TEST DESCRIPTION

The vocabulary of testing is complex and sometimes confusing. It's important to understand the exact nature of these tests in order to appreciate their limitations. Above

all, it's essential to look at the tests themselves, not only at summaries of scores reported in the press or interpreted by school administrators or state departments of education. The popular accounts of the implications of test performance seldom distinguish between reliability and validity, don't describe the distinctions between raw scores and scaled scores and rarely refer back to the test items in detail.

High-stakes tests are used both to determine school careers for individual students and to make judgements about entire schools based aggregated average scores for a school or district. It's important to recognize how the following topics apply to individuals or to groups.

Reliability and Validity

Tests are judged on their reliability and their validity. ***Reliability*** refers to the extent re-testing will produce the same results. ***Validity*** refers to how likely it is that the test assesses what it is intended to assess. An analogy that can be used to understand these concepts is to consider a bullseye target covered with a cloth. If an archer shoots arrows at that target, reliability measures how close the arrows come to each other while validity measures how close the arrows come to the bullseye. Reliability is relatively straightforward and is easily reported in numerical terms. It does not, however, say anything about whether a test provides information that is of any educational value. Validity is much more complex since it refers to the connection between what a test measures and its applicability to real life situations. Some limited forms of validity can be assessed, but, for typical paper-and-pencil tests, they usually refer to the correlation between that test and other, similar tests.

In addition, each of these concepts is applied differently to individuals and to groups. Table 1 provides my personal judgement of the relative reliability and validity of current high stakes tests.

Figure 1
Test Reliability and Validity

	Validity*	Reliability
Individual	Very Poor	Moderate
Large Group	Poor	Good

***** Definition of validity is open to many interpretations. In general, current high-stakes tests are validated by comparison with other tests.

The reliability of large-scale, standardized tests is relatively good for groups of students. It is more doubtful for individual students. We all know cases of children who obtain significant differences in scores on re-testing. Validity depends dramatically on the standard that is applied. For example, SAT tests, used for college admission correlate reasonably well with grades students receive in their first year college courses. They are significantly less valid for predicting college graduation rates.

Current tests

Current tests are supposed to be much improved over the previous generation of tests, which received considerable criticism in the 1960's and 1970's. Yet, they look much like the previous nationally normed, "standardized," multiple-choice tests. Test construction, i.e. developing actual questions, trying them out and refining them, is a highly empirical process. Despite decades of research on alternative test formats, the introduction of performance measures in several states and considerable work by cognitive psychologists on learning, tests have changed very little. For science, this means that they do not actually have the students ***do science,*** but only answer written questions about it, and that they emphasize memorization, recall of facts, vocabulary and, above all, reading ability.

For example, one question on the Massachusetts 2001 5th grade science test[3] asks students to place in appropriate order pictures of tomato plants at various stages of their life cycle. The drawings of the plants themselves are small and, for my older eyes at least, difficult to distinguish, but the drawings also include descriptions identifying them as "tomato plant with flowers," "tomato plant with fruit," "dying tomato plant," "seedling," and "young tomato plant." Any child who can read those phrases has a distinct advantage

over a child who may be a proficient gardener, but has difficulty reading English.

Another question from the same test asks students to interpret a troubleshooting guide for an unidentified machine. The prompt (the text that needs to be interpreted to answer the question) consists of a grid that describes three possible problems for an unidentified appliance and gives various causes and solutions to each. The question then asks the student "According to the guide, what should you do if the machine produces mist but the fan does not work?" The entire question contains 130 words, a complex visual layout and, to the three adults who discussed it to me, such a bewildering array of information that we were not clear about the correct answer after discussing it for several minutes. My point is that this is primarily a reading and vocabulary question, not one that tests the students' understanding of "Technology/Engineering/Engineering Design," as claimed by the reference back to the state standards provided for this question on the Department of Education's web site. As is usually the case in these tests, the two multiple-choice questions described above count equally towards a student's total raw score.

Norming

In principle, one major distinction between the newer tests and the previous generation is that current tests are no longer "normed" but are "criterion referenced." That is, no longer are the test results scored so that all students fall on a normal (Gaussian) distribution curve, with exactly half the students receiving scores in the lower 50% range and the other half receiving scores in the upper 50% of the range.[4] Instead, teams of "experts analyze the questions and performance levels are assigned to ranges of scores. Thus, for the MCAS 2000 test scores were analyzed to determine four bands labeled "Failing," ""Needs Improvement," "Proficient," and Advanced."[5,6].

But norming to some extent is still part of the standardization process, although just how it's applied is often hard to figure out. So that tests results from one year to another can be easily compared, the results are not only equated as described above, but raw scores on tests are also converted to "standard" scores applying complex statistical processes

that alter the distribution of student scores from the simple relationships provided by the raw scores. For example, on the MCAS, all scores are converted to a scale of 200 to 280 regardless of the number of questions on the test.[7] A partial table of the conversion of 4th grade scores for the 2000 Science test is shown in Figure 2.

Figure 2
Standardized Scores, MCAS 4th Grade Science Test, 2000

Raw Score	% of total score	Stand. Score (SS)	% of total SS
0	0%	200	0%
10	19%	204	5%
20	37%	220	25%
27	50%	230	38%
40	74%	250	63%
50	93%	276	95%
54	100%	280	100%

Total points = 54

In order to get a score of 204, the lowest score possible after 200, the student had to get a raw score of 10 (19% of the total points). A raw score of 27 (50% of the possible total score) only yielded a standard score of 230 (38% of the highest possible standard score of 280.) This scoring pattern suggests that the scores are normed at least to some extent. As indicated above, the standardization process was modified for the 2001 MCAS tests. (See endnote 5.)

Besides the uncertainty introduced from the complex methods used to establish standard scores, there is the added problem of the questionable validity of applying normal distributions to measurement of human mental qualities. There is little evidence for the validity of this outside the world of testing. Statistics texts frequently warn against the misuse of normal distribution assumptions beyond their application to a set of random, independent events. Use of such quantitative measures in education was challenged decades ago in two issues of the Elementary Principal, the publication of the National Association of School Principals.[8] That volume included an article by Philip

Morrison entitled, ""The Bell-shaped Pitfall," in which he pointed out that many common properties in nature do not follow a Gaussian distribution.

Test question type

New tests are often described as improved over these of the past because they now contain a wider range of question types. Besides the traditional multiple-choice questions, they now include open-ended questions that require students to supply answers in detail and not only chose among a set of answers given on the test. However, on examining the tests, it appears that many of them are still primarily multiple-choice tests. In a recent review of high-stakes tests in Education Week, the authors conclude

> The Education Week survey also found that while state assessment systems no longer rely solely on multiple-choice questions, they often lack a rich mix of testing formats. Most states use multiple-choice and short-answer items.[9]

Changing tests

Finally, the tests may change from year to year. It's important to keep track of these changes because they profoundly influence the comparisons of scores. For the elementary school science MCAS, the changes from 2000 to 2001, as illustrated in Figure 3 are particularly striking.

Figure 3
2000 and 2001 MCAS Elementary Science Tests Compared

	2000	2001
Grade	4th	5th
Questions	39	20
Total Points	54	26
Multiple-Choice Questions	34	18
Open Response Questions	5	2
Content Strands	4*	4*

*2000 Strands:

- Inquiry
- Domains of Science
  - Physical Science
  - Life Sciences
  - Earth and Space Sciences
- Technology
- Science, Technology, and Human affairs

*2001 Strands:

Strand 1: Earth and Space Science
Strand 2: Life Science (Biology)
Strand 3: Physical Sciences (Physics and Chemistry)
Strand 4: Technology/Engineering

The 2001 test, compared to the previous year's test, is given in a different grade, there are only half as many points and about half as many questions and even the content tested (the standards on which the test is supposedly based) has been modified significantly.[10]

Test Quality

Finally, the tests simply aren't very good — even when assessed in their own terms, accepting the limitations of testing theory. In the past few decades, there have been several extensive studies of the technical and educational quality of large scale, standardized tests. The most thorough was carried out by the Center for Studies in Evaluation at UCLA 25 years ago. The Center developed criteria under four major headings— measurement validity, examinee appropriateness, administrative usability and normed technical excellence—and ranked each of the hundreds of tests they examined as excellent, fair or poor on each of these categories. Of the fifty-eight science tests examined, none received a ranking of excellent on all four categories, and only three were ranked as high as fair on all four. The rest were rated poor in at least one, if not more, of the four possible categories.[11]

More recently, researchers at the Center for the Study of Testing, Evaluation, and Educational Policy at Boston College came to a similar conclusion in a study described by the authors as "the first comprehensive nationwide study to examine commonly used tests, their influence on instruction and the implications for the improvement of math and science instruction. They report that

> [Current tests] fall far short of the current standards recommended by math and science curriculum experts .. . The tests most commonly taken by students—both standardized tests and textbook tests—emphasize and mutually reinforce low-level thinking and knowledge." [12]

The current situation appears to be unchanged, although the newer tests have not been examined as extensively as the previous ones. In the Education Week study mentioned above, the authors report

> All 50 states now have student-testing programs, although the details of such programs vary greatly. Few states, however, have constructed testing programs that adequately measure student achievement against state standards. States often claim their tests are linked to their standards, but research suggests that alignment is not as close as it should be. . . [13]

The National science Foundation has recently awarded AAAS Project 2001 a major ($2.4 million) grant to "develop new strategies and tools for evaluating the alignment of K-12 assessments in science and mathematics to national, state and district standards and benchmarks." [14] Project 2001 staff recognize that current alignment between the tests and the standards is poor and they hope that their work "will result in marked improvement in assessments by influencing tests developers and their customers in the schools." [14]

In summary, standardized, large-scale tests are a very inefficient and crude method for determining individual student achievement. Technically, the tests are not sophisticated and range from mediocre to poor. For assessing overall, system-side achievement they are of limited value, although they can detect major trends. Interpretation of these trends is complex and seldom reported accurately, since there is no simple correlation between all the components—teaching, curriculum, socio-economic factors, school climate—that may effect test results.

Developing Science Assessments

From what we know about teaching and learning, in order to find out what individual students know, we need to give them a chance to express their ideas and then examine these carefully. The rich literature on student conceptions in science, for example, does not use paper and pencil tests, but probes to examine what students know and understand.[15]

The interpretation of student responses is complex and depends on many factors, including precisely how a question is asked. For example, the 2001 MCAS elementary science test includes the following open response question:

Mark has three small rocks about the same size. He wants to know which one is the heaviest but he does not have a scale. Mark has a meter stick, a spring, two baskets with hooks, a pair of scissors, and some string. Explain how Mark could use some or all of these materials to find our which object is heaviest."

The question also shows pictures of the items with the instruction "Use the pictures below to answer question 20."

Research on similar questions that ask students to devise experiments has demonstrated that responses are different if students actually manipulate materials as opposed to simply seeing a picture of a situation. For example, in studying 13 year-olds' answers to a question about how to measure the absorbability of different brands of paper towels, British researchers found that 77% students used some form of quantitative measurement when they actually did the problem using beakers, paper towels and water, while 49% described some quantitative measure when answering in writing in response to a pictorial representation.[16]

Thus, the ability to demonstrate experimental design is dependent on the way a question is asked. The MCAS question, according to the state's analysis, claims to assess knowledge of a different standard; namely "Technology/Engineering/Materials and Tools."

C. POLITICAL REALITIES OF HIGH-STAKES TESTING

But the use or misuse of high-states tests depends more on political considerations than it does on rational arguments about their merits. The tests are here and informal science educators need to address their presence and the conditions that surround their use. Some important considerations are the following:

1. We know that teachers and schools have not accepted (either in UK or USA) better forms of assessment and the testing system always regresses to some form of simple-

to-score questions, with a high percentage of multiple choice items. In Britain, Paul Black and his colleagues recommended a system they called "moderation, "—a linkage of a range of assessment practices and professional development—in a report commissioned by the government.[17] It was never even considered by the ministry to whom it was submitted. In the United States, California developed an elaborate performance assessment system a decade ago, which was tried, but then rejected.

2. Where high stakes tests exist, The tests drive the school day, not the curriculum, standards or policies. There is little sense in expending a great deal of energy in trying to match museum education programs to the state standards when a state has adopted high stakes testing, because it's the tests that matter, whether they conform to the standards and benchmarks or not. It may be useful for political purposes to demonstrate that museum education programs match state standards[18], but that may be irrelevant to the possibility that these programs impact test scores.

3. In the world of high stakes testing, science is not important. Although the movement towards increased testing and attaching high stakes to test results is growing, it's not clear whether this will lead to increased interest in schools on teaching science. First, many states do not include science in the subjects that students need to master in order to graduate. Figure 4 lists the current status of high stakes testing.

Figure 4
Current Status of State Testing[19]

Current Testing States	50
Current "High Stakes" Testing States	22
Current "High Stakes" Science Testing States*	9

*AL, DL, GA, LA, MD, NM, NY, OH VA

Secondly, the political battle over tests usually centers on scores on reading and mathematics tests. As a result school administrators stress language arts and, to a lesser extent, mathematics when they become concerned about test scores. Science tends to get left out. A useful question for teachers in self contained classrooms is to ask them when science is taught during the school day. It is a rare elementary school that

teaches any science in the prime morning hours. In a recent study of 1000 teachers in Colorado to assess the impact of high stakes testing on instruction a decrease in science instruction was one of the major findings.

> Teachers reported either eliminating or cutting back on the amount of science and social studies they teach. Across grade levels, teachers reported reducing the amount of time they spent teaching science and social studies. In some cases, this was because school and district officials had advised them to do so. In other cases it was a choice made by individual teachers.[20]

D. LESSONS FOR SCIENCE MUSEUMS

If we accept the argument that poor quality high stakes tests are here to stay, that they only sometimes include science and that teachers beg science museum educators to help them improve their test scores, what are useful lessons for informal science educators from this situation?

1) Know the "system"

It's important that in all your actions you know at what level you're addressing the educational system, and that you recognize the actions that are appropriate for that level. In your educational work, be aware at what level you interface with the complex multi-layered system that comprises the public schools. It's useful to engage teachers in a conversation about the items on the high stakes tests, but it makes no sense to suggest to them that they take responsibility to change them. It's useful to join committees that are addressing modification of the tests, but only if these are going to coordinate with and have some power to influence the state Department of Education. In your various roles as museum educators, be clear where are you interfacing with the system, and what you can do at that level.

2) Remain engaged

I agree with Susan Sprague, who said earlier in this session,

> Have your museum provide special, specific assistance to districts and consortiums who are working on systemic reform. You can be a major or minor player - but you should be on the team.

It's crucial that science museum educators remain engaged in the reform process, that

is, in the political battles about science education. You have a unique voice, public credibility and resources not available to schools. Use them.

3) Educate yourself

It's important to educate yourself and the people with whom you work about the actual content of the relevant documents, especially the student tests.

What do the national education standards state and what do they imply? National standards have wonderful statements about process, inquiry, etc., but what is picked up most frequently for assessment purposes is the content material that leads to a much more deterministic approach.

You need to become informed and stay informed as the situation changes. Always remember that you are learning about a political issue and its application, not about research.

4) Collaborate

It's crucial that you talk directly with the teachers who bring their classes to your museum. First, you need to get in touch with them just so they know what you have to offer and how you can work together. I've been working with a group of educators at the Museum of Science to look at their school programs and how they relate to school needs and desires. The group has been reminded after doing some interviews with teachers, of how difficult communication with schools and teachers is. There is a constant need to share information, develop programs collaboratively and stay in touch with your clients at the individual teacher level.

5) Modify and update

The consequence of communication and collaboration is that you may frequently have to modify programs and negotiate to keep true to what you think museum visits can accomplish for children and what teachers expect. Unfortunately, the educational scene changes so rapidly, that last year's school program or exhibit based activity may not be

appropriate this year. You have to constantly upgrade and revise what you offer to schools.

6) Be honest

It's terribly important that you make no false promises and only do what you can**.** You need to remind yourself, and tell your clients, the schoolteachers and administrators that SCIENCE MUSEUM EDUCATION PROGRAMS WILL NOT IMPROVE TEST SCORES. [21]The only exception to this generalization would come about if you deliberately modify your programs so they teach test taking skills.

It's certainly possible that students who have come to your program have higher scores afterwards, but never, never even suggest that your program might have anything to do with that. There is almost zero likelihood that there is any correlation between museum programs and test scores. If you take credit for any positive change you will be blamed for a negative change when the inevitable consequences of random distributions comes around.

There are many reasons why museum education won't change test scores. One is that it is actually difficult to demonstrate that ***any*** specific school activities change them very much. In general, standardized tests are good measures of demographic factors, primarily of socio-economic indicators.

In a study of the 1999 MCAS scores, Robert Goudet argues that 86% of the variance in average district scores can be accounted for by six demographic variables: median level of educational attainment, median income level, percentage of households above the poverty level, percentage of single-parent families, percentage of non-English speaking households and level of private school enrollment. [22]

If the entire in-school experience only accounts for 14-16% of the variance in scores, how much can a few hours out of the 180-day school year, influence the scores?

7) Document

Finally, back to evaluation, document what you do and its consequences. One lesson we've learned from the work of the past decades is that although local, detailed documentation and evaluation of the wonderful results of a program are no guarantee that it will be valued by parents, teachers or school systems, the lack of any evidence leaves you particularly vulnerable and defenseless in the face of criticism. If you can show the splendid work the children have done, either in school or in museums, you may get and maintain support.

It is never certain that any evidence is sufficient in the real world. (How much evidence do we need that cigarettes cause cancer?) Likewise, how much evidence do we need that students learn in museums? If you are facing critics with a completely opposed ideology, people who fundamentally believe that children should not be taught to question and investigate, then your evidence for the value of investigations will not sway them. You need to realize that your collect evidence and document your wonderful work for those who support you. They need to be reassured.

---

[1] There is extensive literature supporting this conclusion. A good summary of the misuse of quantitative measures of human qualities to make judgements can be found in Gould. S. J. (1981) ***The Mismeasure of Man***, New York: Norton.

[2] FairTest a Massachusetts based public interest group has been monitoring tests for over two decades, constantly reminding us of the biases in the tests and their frequent misapplication. Their material can be accessed at www.fairtest.org. A leading critic of tests Gerald Bracey, frequently points out the shoddy science involved in test construction, administration and application. For a recent review, see Bracey. E. W. (2001) "The Condition of Public Education," ***Phi Delta Kappan*** 83[2] 157-169.

[3] All the examples in this talk are from the Massachusetts Comprehensive Assessment System (MCAS,) produced by a commercial publisher under contract with the Department of Education, the tests with which I am most familiar. Massachusetts items can be found on the Department of Education web site at http://www.doe.mass.edu/mcas/. Readers from other states are urged to look at items from their tests if these are available.

[4] In a famous study of state reporting of student outcomes on tests used nationally, one inquisitive researcher discovered that all 50 states claimed that their average state scores exceeded 50%. This phenomenon has been termed the Lake Wobegon effect.

[5] For the MCAS 2000 4th grade science test, the threshold maximum score for that level were: Needs Improvement /Failing = 17; Proficient/Needs Improvement = 30, Advanced/Proficient =

43, out of a possible maximum score of 54. The actual setting of these thresholds was based on the scores for the 1998 test by "panelists [who] examined student work in a standard setting process." The 2000 tests thresholds were then determined by a process called "equating," based on average student scores on questions that are the same or comparable on the 1998, 1999 and 2000 tests. The Department recognizes some of the difficulties in the equating process since multiple variables change year to year. "The standards established for these subject areas in 1998 remain unchanged for the 1999 and 2000 MCAS tests. However, because the common test items on the 2000 MCAS tests are different from the common items on the 1998 test, the threshold scores representing those standards may change. In part, a change in threshold scores may be due to differences in difficulty between the 1998 and 2000 MCAS test items. To some extent, however, the change in threshold scores is simply due to the changes in the number of multiple-choice and open-response items and the change in the maximum possible score on most tests. Using the matrix-sampled items that remained unchanged from 1998 to 1999, and from 1999 to 2000, the 1998, 1999 and 2000 MCAS tests are linked through a process called equating. Based on the equating process, the adjustments in the threshold scores needed to maintain the 1998 standards on the 2000 MCAS tests are determined. (THE MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM: Guide to Interpreting the 2000 MCAS Reports for Schools and Districts, p 35)

[6] The actual meaning of any particular score is further complicated if the state changes its scoring rules and thresholds arbitrarily, as Massachusetts has done for the 2001 tests ("State officials alter scoring of MCAS tests," Boston Globe, October 3, 2001.

[7] For some reason all test results for individual students come out to be even numbers

[8] Paul L. Houts, editor, ***The Myth of Measurability***. 1977, New York: Hart. (Reprint of articles from ***The Principal***, 1975,)

[9] Quality Counts, 2001, A Better Balance: Standards, Tests and the Tools to Succeed, Education Week, Vol. 20, No. 17, p. 8-9 (2001) available at: www.edweek.com/sreports/qc01/

[10] Since this talk was delivered, the State Department of Education has announced 2001 MCAS results showing a dramatic increase in scores. Officials indicate that this is the result of successful education reform and that it proves that schools have gotten better. There is little mention of the many possible alternative explanations ranging from simple statistical fluctuations—there is no standard for estimating these and no error ranges are included in reporting average scores—to changes in the tests or changes in the number of students exempted or any of the myriad other factors that can influence test results.

[11] Hoepfner, R., et al. (1976) ***CSE Elementary School Test Evaluations***, Los Angeles, CA: UCLA Center for the Study of Evaluation

[12] Madeus, G. F., et al. (1992) ***The Influence of Testing on Teaching Math and Science in Grades 4-12***, Boston, MA: Boston College. Center for the Study of Testing, Evaluation, and Educational Policy.

[13] Quality Counts, 2001, A Better Balance: Standards, Tests and the Tools to Succeed, Education Week, Vol. 20, No. 17, p. 8-9 (2001) available at: www.edweek.com/sreports/qc01/

[14] AAAS, "Putting Tests to the Test in ***2061 today***, Spring/Summer 2001, Vol. 11 No. 1.

[15] See the description of research on learning in Bransford, J. D., Brown, A. L., and R. D. Cocking, editors, ***How People Learn: Brain Mind, Experience, and School***, Washington, DC: National academy Press. The authors also criticize current assessment practices: "Many standardized tests that are used for accountability still overemphasize memory for isolated facts and procedures, yet teachers are of the judged by how well their students do on such tests." (p. 129)

[16] A review of the factors that influence student responses is found in Murphy, P. "What has

Been Learnt About Assessment from the Work of the APU Science Project?" in Hein, G. E. (editor) (1990) ***The Assessment of Hands-on Elementary Science Programs,*** Grand Forks, ND: North Dakota Study Group on Evaluation.

[17]***Task Group on Assessment and Testing: A Report*** (1987) London: Department of Education and Science,

[18] See "Learning Standards and the Role of Science Centers," ASTC Annual Meeting Panel, October 6, 2001, Jeff Liverman, chair.

[19] www.edweek.com/sreports/qc01/

[20] Taylor, G., Shepard, L, Kinner, F. and J. Rosenthal, (2001) ***A Survey of Teachers' Perspectives on High-Stakes Testing in Colorado: What Gets Taught, What Gets Lost***, Draft Report, CRESST/CREDE/University of Colorado at Boulder, http://education.colorado.edu/EPIC/EPIC%20Documents/COsurvey_draftfinal.pdf

[21] The only exception to this generalization would come about if you deliberately modify your programs so they teach test taking skills.

[22] A more recent study of a sample of Massachusetts communities comes to a strikingly similar conclusion that community income alone is strongly correlated with test scores and accounts for 84% percent of the variance in average scores. Bolon, C. (2001, October 16). "Significance of Test-based Ratings for Metropolitan Boston Schools." Education ***Policy Analysis Archives***, 9(42). Retrieved 10/21/01 from http://epaa.asu.edu/epaa/v9n42/.