The Experimental Design Ability Test (EDAT)

Karen Sirum and Jennifer Humburg

Dept. of Biological Sciences, 202 Life Sciences Building, Bowling Green State University, Bowling Green, OH 43403

Email: ksirum@bgsu.edu

Abstract: Higher education goals include helping students develop evidence based reasoning skills; therefore, scientific thinking skills such as those required to understand the design of a basic experiment are important. The Experimental Design Ability Test (EDAT) measures students' understanding of the criteria for good experimental design through their open-ended response to a prompt grounded in an everyday life science problem. Using a straightforward scoring rubric to analyze student responses, the EDAT provides for consistent and rapid evaluation. Minimal student and classroom time is required to administer the EDAT and it can be used in a pre-/posttest format to measure gains. Significantly, the EDAT is content and terminology independent, and requires minimal quantitative skills. Our findings indicate that the EDAT is sensitive to improvements in experimental design ability, as only students in our sample who participated in a redesigned introductory biology course that included explicit instruction and experiences using the scientific method, made significant gains in their experimental design ability.

Key Words: Scientific Thinking Skills, Science Reasoning, Experimental Design, Assessment, Pretest and Posttest

INTRODUCTION

At the national level, science organizations have expressed their support for science education initiatives that promote scientific literacy (American Association for the Advancement of Science, 1989 and 2011; National Research Council, 1995; National Science Foundation, 1996; Osborne, 2010). A scientifically literate person is one who is able to evaluate the quality of scientific information, pose and evaluate arguments based on facts, and apply this information appropriately (National Research Council, 1996). We set out to design an assessment instrument that would allow us to determine whether we were providing such a learning environment for undergraduate non-science majors in a redesigned introductory biology course. We devised an assessment instrument called the Experimental Design Ability Test (EDAT) and we investigated the test's sensitivity by evaluating students' ability to design an experiment at the beginning and end of the course.

The EDAT requires that students explain how they would go about determining whether they would accept a claim about a product in an open-ended question format. First students have to recognize that an experiment *can* be done to evaluate the plausibility of the specified claim. Then they guide us through their thinking process in the design of such an experiment. Students need to demonstrate their understanding of the importance of controlling variables, larger sample sizes, reproducibility, and of the limitations to the generalization of their conclusions. However, the EDAT is content independent and does not require students to use specific terms such as independent or dependent variables and it has a minimal requirement for quantitative skills. Compared to a multiple-choice test, this format gives insight into a student's thought processes instead of simply the end result of their thinking. It demands that students think through the process of designing an experiment in their own minds without being cued in on what the correct answer might be from the options provided in a multiple-choice test (Lederman, 1998). The EDAT only requires 10-12 minutes for students to complete, and scoring is straightforward using our specific scoring rubric, requiring one hour for the instructor to score 40 tests.

METHODS

The Experimental Design Ability Test (EDAT) was administered in a pre- and posttest format to students enrolled in multiple sections of a non-majors introductory biology course. Sections of the course, including those using non-lecture based teaching strategies and student-designed labs (Experimental Groups) and those using traditional lecture and descriptive labs (Traditional Groups), were assessed with the EDAT. In Experimental Groups, interactive engagement teaching strategies were used which involved challenging students with a variety of problem based, interactive, and group learning activities, and incorporating Socratic discussions (Klionsky, 2004; Knight and Wood, 2005). In addition, in Experimental Groups the traditional lab component was replaced by lab activities that involved student-designed experiments, some based

Pretest: *Advertisements for an herbal product, ginseng, claim that it promotes endurance. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.*
Posttest: *The claim has been made that women may be able to achieve significant improvements in memory by taking iron supplements. To determine if the claim is fraudulent and prior to accepting this claim, what type of evidence would you like to see? Provide details of an investigative design.*

Fig. 1. EDAT pre- and posttest student prompts. Students are provided a sheet of paper with the prompt at the top and told to use as much writing space and time as they need.

on the "desk-top" biology labs of Handelsman *et al.* (2002). These sections of the course met 3 times a week: once a week for 2 hours and twice a week for 75 minutes. The maximum enrollment was 25 students. In Experimental Groups 1-3, the instructor integrated lab group learning activities, discussions, and group problem solving, while lecturing was limited to approximately 15 minutes maximum per class. Students completed web-based readings and take home reading quizzes to prepare for class (Klionsky, 2004). In contrast, Experimental Group 4 fully incorporated the same lab experiences as Experimental Groups 1-3 (see below) but the main pedagogical strategy was lecturing with occasional (approximately once per week) implementation of some of the active learning activities used in Experimental Groups 1-3.

The lab activities for students in Experimental Groups 1-4 involve first presenting students with a problem and some background information. Students are then asked to propose a hypothesis and design an experiment to test their hypothesis regarding the problem in a PreLab homework assignment (North Carolina State University, 2004). Two such examples of these types of lab experiences are the "Moldy Bread" and "Mutation and Selection" lab activities found in Handelsman *et al.* (2002). In groups and with the entire class, students discuss, critique, and modify their experimental design, and then perform their experiments in pairs. Individually, students submit a written lab report, using as a guide a lab report rubric that we abbreviated and modified based on that described in LabWrite (North Carolina State University, 2004; Appendix). These lab reports, a maximum of 3 pages including figures and tables, include a student's hypothesis, methods with description of treatment and control groups and variables, results, discussion, brief conclusions, and references. The lab report is then critiqued by the instructor and returned to students with comments and with a marked copy of the rubric, indicating the student's level of achievement and the number of points earned in each section of the lab report. Students complete six PreLab assignments, perform six experiments and write six lab reports throughout the semester, in addition to other activities. Through the student-designed experiments, an emphasis is placed on helping students learn to be precise in their interpretation of their data and thorough in their explanation of the limitations of their experimental designs and conclusions. A student's combined average on these lab assignments comprises one third of their overall course grade. These groups are referred to as "Experimental" because they utilized new science teaching strategies for our institution.

In the Traditional sections of the course, students attended 50-minute lectures three times a week and

Table 1. Characteristics of the introductory non-majors biology course sections in this study.

	Section Characteristics							
Section	Term	# Enrolled*	N EDAT	% Freshmen	% Female	Instructor**	Lab Method†	Lecture Method†
Exp. 1	Sp '07	21	21	57%	57%	A	S-D	AL
Exp. 2	F '07	25	24	100%	80%	A	S-D	AL
Exp. 3	F '07	22	21	23%	64%	A	S-D	AL
Exp. 4	F '07	24	22	100%	58%	B	S-D	Trad. + AL
Trad. 1	F '07	15	12	8%	62%	C	Trad.	Trad.
Trad. 2	F '07	21	20	100%	56%	C	Trad.	Trad.
Trad. 3	F '07	123	76	40%	66%	C	Trad.	Trad.
Trad. 4	Sp '08	119	71	48%	52%	B	Trad.	Trad.

Exp. = Experimental Groups. Trad. = Traditional Groups.
% Freshman & % Female are tabulated from the sample of students that participated in assessments.
Term is the year of either the Spring (Sp) or Fall (F) semester for that course section.
*Indicates number of students enrolled in course, which is different from sample number, N.
**Course sections were taught by three different instructors indicated by A, B, or C.
†Lab teaching strategy is either S-D (Student-Designed labs) or Trad. (Traditional descriptive lab). Lecture teaching strategy is AL (Active Learning/Interactive Engagement), Trad. (Traditional lecture), or Trad. + AL (a combination of both).

participated in traditional descriptive labs once a week for a 2- hour time period. In the Traditional lab, students receive credit for completing lab worksheets, tests and quizzes, and an end-of-semester PowerPoint presentation based on one of their lab activities. A student's scores on these assessments comprise one third of their course grade. Two of the Traditional sections of the class had enrollments similar to the Experimental Groups (maximum 25 students), while two of the Traditional sections were large lecture sections with up to 140 students and with multiple 30-student sections for the lab taught by various graduate Teaching Assistants (See Table 1).

The EDAT pretest was administered during the first week of the semester in each of the participating sections. Students were given as much time as they needed to complete their responses; almost all students finished in 10-12 minutes. Grade points were not awarded to the students for participation in the pretest. We have found that students are often eager to do their best at the beginning of the semester and motivation is high. Students were informed that we were gauging their abilities in biology to gain a better understanding of where they were and how we could help them be successful in this course. EDAT pretest scores and the EDAT scoring rubric were not shared with students at any time and students were not told in advance that a similar posttest would be administered later in the semester.

The EDAT posttest was administered during the week prior to the last week of class in the Experimental Groups. Students were told several days in advance that they were having a quiz based on the scientific thinking skills they learned and that their effort on the quiz would count for approximately 3% of a student's grade in the course. In the Traditional Groups, the EDAT posttest was also administered during the week prior to final exams. Students were also told in advance that they were having a quiz based on scientific thinking skills and that they could earn bonus points towards their course grade (not more than 1% of the course total) for their degree of effort on the EDAT. Given the different instructors and course formats, we were not able to control the weight given to the EDAT towards a student's grade in the different sections of the course. Again, students were given as much time as they needed to complete their responses, typically 10-12 minutes. Note that all three of the course instructors and the laboratory Teaching Assistants had equal knowledge of the EDAT pre- and posttest prompts and scoring rubric.

Table 2. Determination of inter-rater reliability value for EDAT scores: Pearson's coefficient.

EDAT Inter-rater Reliability		
	$r = 0.835$	$p < 0.001$
	M	*SD*
Rater 1	5.16	2.54
Rater 2	5.56	2.31

Only EDAT scores for students that participated in both the pre- and posttest are reported and were used in statistical analysis. The data did not fit a normal distribution; therefore, a nonparametric test, the One-sample Wilcoxon sign rank test, was used for analysis. Individual student scores were paired in the Wilcoxon test. Statistical analysis of results was performed using Minitab 15 Statistical Software (2008). Correlation analyses (Spearman's rank correlation) were performed using STATISTICA (2008).

RESULTS AND DISCUSSION

Criteria for a Scientific Thinking Test

We wanted to measure changes in students' scientific thinking in terms of their ability to design experiments, and we wanted to be able to use an assessment instrument with the following six criteria:

1. Not time consuming to administer to students in the classroom,
2. Based on a practical challenge from an "everyday life" problem to increase student buy-in and effort,
3. Requiring minimal student quantitative skills,
3. Open-ended to reveal student's thinking (*i.e.*, not multiple choice),
4. Easy to score consistently, and
6. Providing a quantitative measure.

Therefore, we designed the Experimental Design Ability Test (EDAT). Students were given a specific prompt asking them to come up with an investigative design to test a claim (based on Ommundsen, 2005; Fig. 1). We chose the wording of the prompt to avoid leading the students or directing their response as much as possible and we avoided using multiple-choice questions because they may provide unintended corrective feedback to the students.

The open-ended student responses were then scored using a rubric that we designed for the purpose of simplification and clarification of the criteria for a good experimental design (Allen and Tanner, 2006; Moskal, 2000; North Carolina State University, 2004; University of Michigan-Dearborn School of Education, 2002). Ten criteria were selected for good basic experimental design (Fig. 2) and each student's score reflects the number of criteria correctly included in their answer, with a maximum score of 10. Note that the order of the listed criteria was designed to reflect increasingly difficult items for students to include in their EDAT response such that, for example, the tenth point is more challenging for the student to include than the first point. Work is in progress to confirm this order

EDAT Scoring Rubric (7/2010)

____ 1. Recognition that an experiment can be done to test the claim (vs. simply reading the product label).
____ 2. Identification of what variable is manipulated (independent variable is ginseng vs. something else).
____ 3. Identification of what variable is measured (dependent variable is endurance vs. something else).
____ 4. Description of how dependent variable is measured (e.g., how far subjects run will be measure of endurance).
____ 5. Realization that there is <u>one other</u> variable that must be held constant (vs. no mention).
____ 6. Understanding of the placebo effect (subjects do not know if they were given ginseng or a sugar pill).
____ 7. Realization that there are <u>many</u> variables that must be held constant (vs. only one or no mention).
____ 8. Understanding that the larger the sample size or # of subjects, the better the data.
____ 9. Understanding that the experiment needs to be repeated.
____ 10. Awareness that one can never <u>prove</u> a hypothesis, that one can never be 100% sure, that there might be another experiment that could be done that would disprove the hypothesis, that there are possible sources of error, that there are limits to generalizing the conclusions (credit for any of these).

Fig. 2. EDAT scoring rubric used to score students responses to the edat prompts. This rubric should <u>not</u> be shared with students. Each item that is included in the student's response is checked and the checks are tallied for a student's total EDAT score with a maximum of 10 points.

(manuscript in preparation). Of significance is the fact that the EDAT scoring rubric allows students to demonstrate understanding of experimental design without having to use any specialized vocabulary or terms such as "independent/dependent variable" or "control", and without requiring substantial quantitative skills. The EDAT can be used with both science majors (manuscript in preparation) and non-majors at all levels.

Using the rubric, students' EDAT responses were independently scored by two raters, and then scores were compared for inter-rater reliability. Each rater was able to score approximately 40 EDAT responses in an hour and the inter rater reliability was determined to have a Pearson's Correlation Coefficient of 0.83 (Table 2). Note that at time of scoring, raters did not know the identity of the course section of each student response.

Changes in Students' Experimental Design Ability

To investigate the utility of the EDAT, we administered it to multiple sections of the same introductory non-majors biology course that focused on the molecular and cellular basis of life at our large undergraduate and graduate degree granting public university (~19,000 undergraduate students) over four 16-week semesters.

Three different teaching strategies were used in various sections of the course, each of which included a 3-credit lecture and 1-credit laboratory component: some sections fully used interactive engagement, non-lecture-based teaching strategies as well as student-designed labs (Experimental Groups 1-3, Table 1); one section had a combination of mainly traditional lecturing with some active learning activities but did fully incorporate the student-designed labs (Experimental Group 4, Table 1); and some sections consisted of traditional lecture and traditional descriptive lab teaching methods (Traditional Groups 1-4, Table 1).

Three different instructors taught these different groups (Table 1). One instructor taught both the lecture and the lab sections of Experimental Groups 1-3. Another instructor taught the Experimental Group 4 lecture sessions and a biology graduate student teaching assistant (TA) who was trained for instruction of the student-designed laboratories, taught the lab sessions of this group. The same instructor for Experimental Group 4 also taught Traditional Group 4 with various other TAs teaching the multiple 30-student lab sections. A third instructor taught the Traditional Groups 1-3, again with various TAs teaching the lab sessions (Table 1). Note that Traditional Group 1 was an honors section of the course restricted to students with a minimum university grade point average (GPA) of 3.5.

It is important to understand that the purpose of this report is NOT to compare two teaching strategies or different instructors *per se*, but rather to use the existing differences in teaching strategies to test the utility and significance of the EDAT. The intent of the Experimental Groups was to help develop students' scientific thinking skills; specific student learning activities were incorporated into both the lecture and lab portion of the course to this end. Therefore, it is reasonable to predict that these sections of the course will show gains in EDAT scores, thus serving as one form of validation for the EDAT instrument. While it is natural to compare the various sections of the course to identify those factors contributing to the differences in EDAT scores, there are confounding variables that limit the interpretation of data to this end. Further analysis of the many course section differences, and analysis of varied teaching strategies used, is the subject of ongoing investigation and will be published separately.

We analyzed the utility of the EDAT by administering it to non-major introductory biology students enrolled in both Experimental and Traditional Groups at the beginning of the 16-week

Table 3. The means and standard deviations of pre- and post-test EDAT scores for each course section.

	Mean EDAT Scores			
	Pretest		Posttest	
Section	*Mean*	*SD*	*Mean*	*SD*
Exp. 1 n = 21	4.67	1.74	6.52	0.93
Exp. 2 n = 24	3.29	1.73	7.21	1.64
Exp. 3 n = 21	3.14	1.91	6.95	1.32
Exp. 4 n = 22	3.33	1.88	5.77	1.51
Trad. 1 n = 12	3.33	1.92	3.50	1.24
Trad. 2 n = 20	3.30	1.76	3.05	1.43
Trad. 3 n = 76	3.00	1.76	3.42	1.81
Trad. 4 n = 71	3.66	2.07	3.61	1.69

semester and again at the end. While the basic format of the EDAT from pretest to posttest does not change, and the requirements to answer the question are the same for both, in this data set, the specific details of the question posed to the student was varied pretest to posttest so that the prompt seemed different to the students (Fig. 1). We have subsequently used the pretest prompt also as the posttest prompt with other classes and observe results similar to that reported here, indicating that the differences in the two prompts is not sufficient to account for differences in EDAT pre- and posttest scores (manuscript in preparation).

After scoring the pre- and posttest EDAT responses, we found that the average EDAT pretest scores for the Experimental and Traditional Groups were 3.6 (SD=1.8) and 3.3 (SD=1.9) respectively (Table 3 & Figure 3) indicating that for both groups, students' experimental design abilities are very similar at the beginning of the semester. The average EDAT posttest scores for the Experimental and Traditional Groups were 6.6 (SD=1.35) and 4.0 (SD=1.5) respectively (Table 3 & Figure 3). Statistical analysis of these EDAT scores indicates that all of the sections that incorporated student-designed experiments in the laboratory sessions (Experimental Groups 1-4) made statistically significant gains ($p<0.001$), while the sections with traditional labs (Traditional Groups 1-4) did not make gains (Table 4) indicating that something about the students' experience in the Experimental Groups facilitated the development of experimental design ability compared to the Traditional Groups. This can also be seen when looking at the distribution of the number of students with scores from 1-10 on the EDAT pre- and posttest for both Experimental and Traditional Groups (Figs. 4 & 5). Students who were exposed to experiences that required them to design their own experiments and analyze and reason with data made greater gains in their ability to design experiments as measured by the EDAT compared to their peers who were exposed to traditional lecture and laboratory teaching methods.

Table 4. Determination of significant gains in EDAT scores for each participating introductory non-majors biology course section: results from Wilcoxon Sign Rank test.

	EDAT Change from Pre- to Posttest			
Section	*Mean*	*SD*	*Median*	*P*
Exp. 1 n = 21	2.28	2.03	2.5	**< 0.001***
Exp. 2 n = 24	3.92	2.28	3.5	**< 0.001***
Exp. 3 n = 21	3.81	2.16	4.0	**< 0.001***
Exp. 4 n = 22	2.55	1.99	2.5	**< 0.001***
Trad. 1 n = 12	0.17	2.25	0.0	0.894
Trad. 2 n = 20	-0.25	1.77	0.0	0.570
Trad. 3 n = 76	0.42	2.09	0.5	0.063
Trad. 4 n = 71	-0.06	1.87	0.0	0.768

$^{*}p<0.05$

Although the goal of the research reported in this paper is not a controlled experiment to assess the effectiveness of an instructional method, the differences in instructional methods used in the Experimental and Traditional Groups gave us an opportunity to assess experimental design ability in students who either were or were not exposed to student-designed experiment experiences. While these data do not rule out other differences between the two groups that may influence performance on the EDAT, the data do show that the EDAT can be used to rate students' scientific thinking ability in terms of their understanding and application of the fundamental concepts of experimental design. Halpern (2003) found similar results when using a standardized thinking skills tests to assess the effectiveness of critical thinking instruction and found that students who were taught with specific thinking instruction outperformed those who were not taught in this manner. Our findings are supportive of the effectiveness of our instrument, the EDAT, in measuring basic experimental design ability, as students who had an opportunity to use and practice experimental design were the same students who made gains in their EDAT scores.

The reliability of the EDAT can be demonstrated by the fact that all of the sections that incorporated student-designed experiments made statistically significant gains in EDAT scores. The validity of this assessment has been established thus far through scoring the EDAT responses with a scoring rubric that consists of a list of the elements that make a good experimental design (Fig. 2), thereby establishing face or qualitative validation. If further validation is necessary, this could be accomplished through the use of other measures of experimental design ability. In this regard, we could not use students' scores on their lab reports as a direct report of their change in experimental design ability since lab reports also require conceptual understanding of

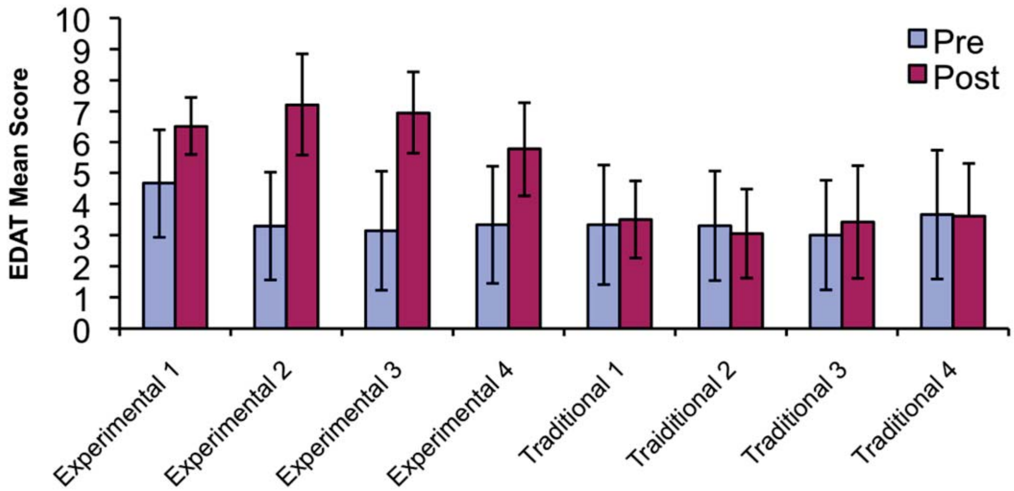

Fig. 3. Pre- and post- EDAT score means +/- standard deviation for each participating introductory non-majors biology course section. Gains in average EDAT scores for the experimental groups are statistically significant (p<0.001) as determined by the one-sample Wilcoxon Sign Rank test. Traditional groups did not make statistically significant gains (See Tables 3 and 4).

the biological concepts being explored. Early labs were conceptually simpler than those the students performed later in the semester so a student's change in lab report scores would not necessarily reflect changes in their experimental design ability.

In our analysis, we had a total of three instructors. One instructor (C in Table 1) taught only Traditional Groups 1-3, another instructor (B, Table 1) taught Experimental Group 4 and Traditional Group 4, and a third instructor (A, Table 1) taught only Experimental Groups 1-3. We observe increases in EDAT scores for two different instructors. Experimental Groups 1-4, with either instructor A or B, all incorporated the student-designed experiments and all made statistically significant gains in the EDAT (p<0.001). Instructor B also taught Traditional Group 4 without the student-designed experiments and this group's average EDAT score virtually remained the same over the course (pre=3.66, post= 3.61; Fig. 3). This example points

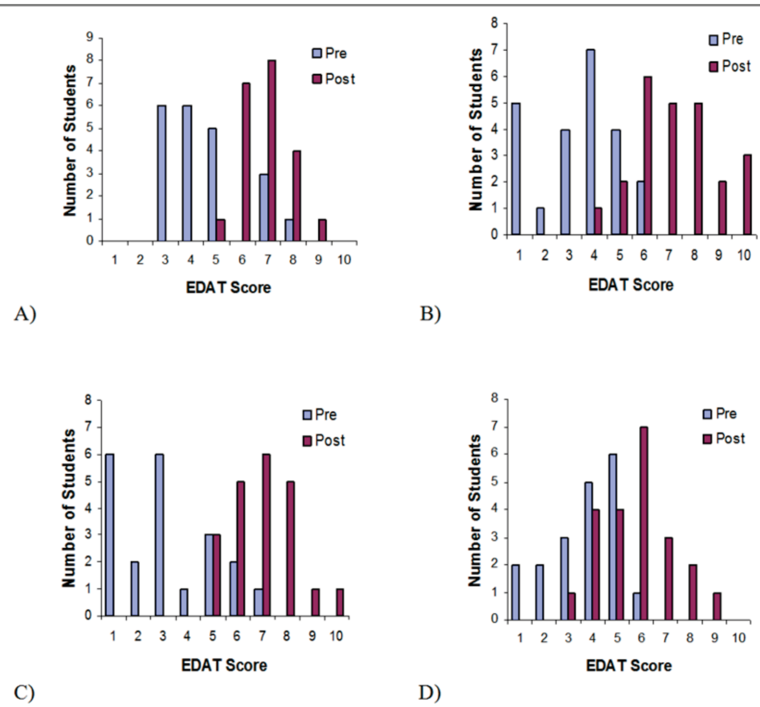

Fig. 4. (A-D) Frequency distribution of student pre- and posttest EDAT scores for participating experimental teaching sections of introductory non-majors biology. The x-axis of the graphs shows the EDAT score and the y-axis shows the number of students with that score.
(A) Experimental Group 1.
(B) Experimental Group 2.
(C) Experimental Group 3.
(D) Experimental Group 4.

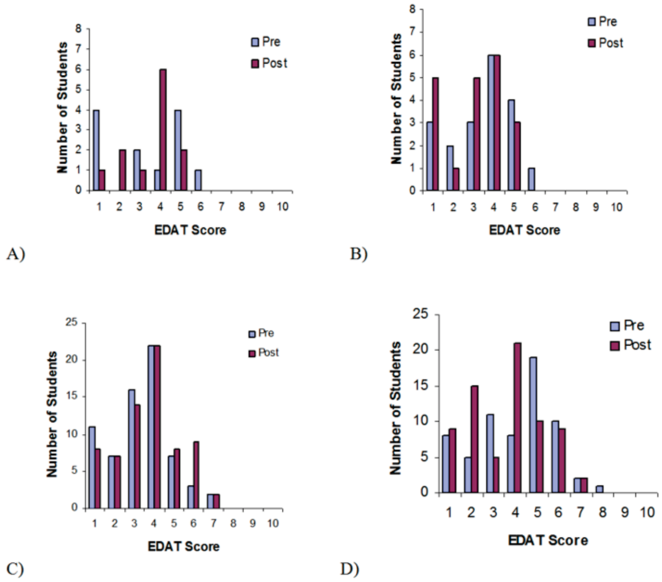

Fig. 5. (A-D) Frequency distribution of student pre- and posttest EDAT scores for participating traditional teaching sections of introductory non-majors biology. The x-axis of the graphs shows the EDAT score and the y-axis shows the number of students with that score.
(A) Traditional Group 1.
(B) Traditional Group 2.
(C) Traditional Group 3.
(D) Traditional Group 4.

out that gains in the EDAT are not limited to one instructor. Since this is a small sample size, further work is in progress using the EDAT in many other courses with other instructors and will help to clarify this issue.

Demographic Factors That Do Not Impact EDAT Scores

We were interested in finding out if other factors influence EDAT scores. With the use of a two-sample t-test, no male-female differences were found in EDAT gains (p=0.961) suggesting that the teaching techniques were similarly beneficial or not for both male and female students and that the EDAT is not biased with regard to gender. To find out if there is a statistically significant difference in EDAT pre-scores or gains depending on age or pre-score (note: in our data set students ranged in age from 18-25 years), we used a One-way ANOVA. The data indicate that the mean scores of students ages 18-25 years do not differ: students did not come into the course with a higher score because of their age, and those students who did make gains did so regardless of their age and pre-score.

One might think that students who have more college experience in general would perform better on the EDAT. Students who have more college experience may have had more science courses or other courses or experiences that promoted the development of their scientific thinking. Using a One-way ANOVA we looked at the difference in pre-scores and gains made between freshman, sophomores, juniors or seniors on the EDAT. Results indicated that there was no difference: on average, students in our sample, regardless of their year in college, are not entering introductory non majors biology with the ability to score above 3.6 on the EDAT. Although we did not find differences between pre-scores or gains for this sample of undergraduate students enrolled in non-majors introductory biology for the EDAT based on gender, age, or year in college, some differences may exist. Current work involves a larger sample of students who also include science majors.

CONCLUSIONS

The EDAT is sensitive to improvements made in experimental design ability: the Experimental Groups made significant gains and the Traditional Groups did not make gains. It is possible that some of the Experimental *vs.* Traditional Group differences in EDAT scores are due to other unexamined factors such as incoming ACT or SAT scores, GPA, or previous science courses the students have taken, so conclusions about the differences in outcomes among these groups are limited. However, Traditional Group 1, an Honors section of the course requiring students to have a minimum 3.5 GPA, did not have an average EDAT score that was different from the non-Honors sections of the course. Current research involves investigating the role of these variables when the sample size is larger and includes science majors, and whether EDAT gains are maintained with time after the end of the course. The main purpose of this study was to design, implement, and determine some of the characteristics and effectiveness of the EDAT, our diagnostic test for experimental design ability, in an introductory biology course, and our data demonstrate the utility of the EDAT to this end. The EDAT was designed to be content independent so that it can be used with any student population. Similarly, by design, the EDAT has a very low requirement for quantitative skills. For example, it is sufficient for students to understand that a large sample size is desirable, however, knowing the actual number of subjects that is sufficiently large for statistically significant data is beyond the scope of the EDAT. Our reasoning for this approach is that we expect that not all students will have highly developed experimental design skills. Rather our goal is that, in everyday life decisions, all students will be aware of the criteria for good experimental design, have the ability to ask questions of data to help them determine if conclusions are warranted, and will understand the limitations to conclusions.

ACKNOWLEDGEMENTS

The authors thank our colleagues for allowing us to test the EDAT in their classrooms and the reviewers for their helpful comments.

HUMAN SUBJECTS

This work is in compliance with the policies and has the approval of the Bowling Green State University Human Subjects Review Board.

REFERENCES

ALLEN, D. AND K. TANNER. 2006. Rubrics: Tools for Making Learning Goals and Evaluation Criteria Explicit for both Teachers and Learner. *CBE-Life Sciences Education* 5: 197-203.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. 1989. *Science for All Americans. A Project 2061 Report on Literacy Goals in Science, Mathematics, and Technology*. AAC&U, Washington, DC.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE. 2011. *Vision and Change in Undergraduate Biology Education: A Call to Action.* AAC&U, Washington, DC. Accessed from http://visionandchange.org/ on 10 April 2011.

HALPERN, D. 2001. Assessing the Effectiveness of Critical Thinking Instruction. *J. General Education* 50(4): 270-286.

HANDLESMAN, J., HOUSER, B., AND H. KRIEGEL. 2002. *Biology Brought to Life: A Guidebook to Teaching Students to Think Like Scientists.* McGraw-Hill, NY.

KLIONSKY, D.J. 2004. Talking biology: learning outside the book—and the lecture. *Cell Biol. Educ.* 3: 204 -211.

KNIGHT, J. K. AND W.B. WOOD. 2005. Teaching More by Lecturing Less *Cell Biol. Educ*. 4: 298-310.

LEDERMAN, N.G., WADE, P.D., AND R.L. BELL. 1998. Assessing the nature of science: What is the nature of our assessments? *Science and Education* 7(6): 595-615.

MINITAB 15 STATISTICAL SOFTWARE. 2008. Minitab Incorporated. State College, PA.

MOSKAL, B. M. 2000. Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation* 7(3). Accessed from http://PAREonline.net/getvn.asp?v=7&n=3 on 14 April 2007.

NATIONAL RESEARCH COUNCIL. 1995. *National Science Education Standards*. National Academy Press, Washington, DC.

NATIONAL RESEARCH COUNCIL. 1996. *National Science Education Standards*. National Academy Press, Washington, DC.

NATIONAL SCIENCE FOUNDATION. 1996. *Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology*. Directorate for Education and Human Resources, Washington, DC.

NORTH CAROLINA STATE UNIVERSITY. 2004. *LabWrite: Improving Lab Reports.* Accessed from http://labwrite.ncsu.edu/ on 10 April 2011.

OMMUNDSEN, P. 2005. Critical Thinking in Biology Using Case Studies. Accessed from http://capewest.ca/ct.html#Judging%20Testability%20of%20a%20Claim on 12 July 2010.

OSBORNE, J. 2010. Arguing to Learn in Science: The Role of Collaborative, Critical Discourse. *Science* 328: 463-466.

STATISTICA (2008). StatSoft, Incorporated. Tulsa, OK. Accessed from http://www.statsoft.com/ on 13 July 2010.

UNIVERSITY OF MICHIGAN-DEARBORN, SCHOOL OF EDUCATION. 2002. Rubric for "Experimental Design" component of the Science Education Portfolio (SEP). Accessed from http://www.umd.umich.edu/sep/rubric_expdes.htm on 12 July 2007.

APPENDIX

EVALUATION: Lab Report ______ Writer: ____________________

Section Points		no credit	attempted	partial credit	full credit	Section Scores
3	**Title**					
	Describes lab content concisely, adequately, appropriately		1	2	3	
15	**PreLab**					
	Effectively defines the research problem and states the research question or goal					
12	**Hypothesis**					
	States hypothesis and provides logical reasoning for it		4	8	12	
12	**Methods/Treatments/Controls**					
	Gives enough details to allow for replication of procedure		2	4	6	
	Clearly identifies treatment(s) and necessary controls		2	4	6	
18	**Results**					
	Opens with effective statement of overall findings		2	4	6	
	Quantifies results if possible		1	2	3	
	Accurately and carefully measures data/makes observations		2	4	6	
	Format of tables and figures is clear and correct		1	2	3	
24	**Discussion**					
	Logically explains why results support or refute hypothesis		2	4	6	
	Backs up statement with reference to specific findings		2	4	6	
	Thoughtfully demonstrates clear understanding of limitiations to conclusions		2	4	6	
	Suggests and describes follow-up experiment or question		2	4	6	
6	**Conclusion**					
	Convincingly describes what student learned from this lab experience		2	4	6	
1	**References**					
	All appropriate sources in the report are listed and citations and references adhere to proper format		x	x	1	
5	**Presentation**					
	Report is written in scientific style: clear and to the point		1	2	3	
	Grammar and spelling are correct		x	x	2	
4	**Overall aims of the report:** *the student...*					
	Has successfully learned what the lab is designed to teach		x	2	4	
100					**Points Earned**	
					Total Possible Points	100

Appendix. Lab report rubric modified from LabWrite (North Carolina State University, 2004) and used with students as a guide and for grading lab reports in the experimental sections of the introductory biology course.