

Assessing Instructional Sensitivity Using the Pre-Post Difference Index: A Nontechnical Tool for Extension Educators

Abstract

This article provides an illustrative description of the pre-post difference index (PPDI), a simple, nontechnical yet robust tool for examining the instructional sensitivity of assessment items. Extension educators often design pretest-posttest instruments to assess the impact of their curricula on participants' knowledge and understanding of the concepts taught. Although the use of pretests and posttests is common in Extension evaluation, the validity and reliability of these tests are rarely reported or discussed, mostly due to many Extension educators' limited knowledge of various statistical methods. The PPDI method described in this article should be a useful addition to Extension educators' evaluation toolboxes.

Keywords: [pre-post difference index](#), [program evaluation](#), [pretest-posttest](#)

Omolola A. Adedokun

Senior Extension
Evaluation Specialist
University of Kentucky
Lexington, Kentucky
Lola.adedokun@uky.edu

Introduction

Evaluation of the effectiveness and impact of Extension programs often involves the use of pretest-posttest designs to assess participants' knowledge and understanding of instructional materials. Although the use of pretests and posttests is common in Extension evaluation, the psychometric properties (e.g., validity and reliability) of these tests are rarely examined or discussed, not because Extension educators undervalue reliable and valid instruments but because many have limited knowledge of evaluation methods (Arnold, 2006). My one-on-one conversations with my Extension colleagues have suggested that Extension educators desire to know whether the tests and instruments they have developed to evaluate their programs are indeed measuring participant learning gains. However, as noted by Arnold (2006), most Extension educators have limited statistical skills, including the technical expertise required to conduct psychometric analyses. Extension educators are also faced with shrinking evaluation resources (Silliman, 2016), including the financial resources needed to hire the services of a statistician or an external evaluator.

Their limited evaluation skills notwithstanding, Extension educators are becoming increasingly interested in adopting rigorous evaluation practices and incorporating evaluation into the program planning process and reporting results (Wise, 2017). Given dwindling federal funding and stricter requirements for accountability and documentation of program impact to justify continued funding (Lamm, Israel, & Diehl, 2013; Silliman, 2016;

Wise, 2017), there is the need for articles and how-to papers that describe simple and nontechnical tools Extension educators can use to enhance the evaluation of their programs.

My goal with this article is to describe a simple tool for assessing the validity of pretest-posttest evaluation instruments. Specifically, I provide a nontechnical illustrative example of how to examine instructional sensitivity (a measure of validity) using the pre-post difference index (PPDI). I also discuss how the results or findings of analysis can be used for improving instruction and adjusting evaluation instruments.

Instructional Sensitivity and the PPDI

Instructional sensitivity is a measure of the validity of assessment instruments. It is defined as the "extent to which the assessment items represent the enacted curriculum" (Li, Ruiz-Primo, & Wills, 2012, p. 3) or the extent to which "students' performances on a test accurately reflect the quality of instruction specifically provided to promote students' mastery of what is being assessed" (Popham, 2007, p. 146). More simplistically, instructional sensitivity refers to the extent to which the questions on a test measure the material or content covered in the curriculum (Lan et al., 2012). That is, it measures "how students react" to the learning materials or intervention (Xie, Zhang, Nourian, Pallant, & Bailey, 2014, p. 761).

The PPDI measures the "pre-post" differences in item difficulty after an instruction or education program has been implemented (Huang et al., 2015) and is mostly useful for the analysis of dichotomous items (e.g., yes-no, correct-incorrect). The method for calculating PPDI is appropriate for examining validity in pretest-posttest designs as well as in intervention versus control (comparison) conditions.

Methodology: How to Calculate the PPDI

For illustration, consider a hypothetical case in which an Extension educator has developed a 10-item pretest-posttest evaluation instrument to assess the impact of a curriculum that was delivered to 120 participants. This educator would calculate the PPDI by following the steps described here.

1. Grade the pretest and the posttest.
2. Calculate the numbers (and percentages) of participants who answered each question correctly at pretest and posttest (see Table 1).

Table 1.

Numbers of Participants Who Provided Correct Responses on a Pretest-Posttest Instrument

Question	Correct responses (pretest)		Correct responses (posttest)	
	<i>f</i>	%	<i>f</i>	%
Question 1	19	15.83	103	85.83
Question 2	26	21.67	98	81.67
Question 3	45	37.50	108	90.00
Question 4	57	47.50	102	85.00

Question 5	58	48.33	95	79.17
Question 6	30	25.00	55	45.83
Question 7	84	70.00	97	80.83
Question 8	100	83.33	105	87.50
Question 9	36	30.00	38	31.67
Question 10	52	43.33	38	31.67

3. Calculate the item difficulty index (IDI) for each question on the pretest and posttest. Item difficulty refers to the proportion of participants who answered a question correctly. For example, 30 of the 120 participants, or 25.00%, answered Question 6 correctly on the pretest, and 55 participants, or 45.83%, answered the question correctly on the posttest. Hence, the difficulty index for this question is 0.25 at pretest and 0.46 at posttest (see the second and third columns of Table 2).
 - In general, IDI ranges from 0 to 1, where 0 indicates a very difficult question that no participant answered correctly and 1 indicates a very easy question that all participants answered correctly. According to Huang et al. (2015), appropriate IDIs range from 0.30 to 0.70. Questions with IDIs less than 0.30 are considered to be very difficult, and questions with IDIs of 0.70 and above are considered to be easy (Huang et al., 2015). Typically, questions are expected to be difficult for students before instruction and easier after instruction.
4. Calculate the PPDI for each item by subtracting the IDI at pretest from the IDI at posttest (see the fourth column of Table 2).
5. Interpret the PPDI for each question (see the last column of Table 2). In general, questions that elicit greater PPDI values are more sensitive to instruction (i.e., better able to measure student learning). Li et al. (2012) employed the following rules of interpretation:
 - If the PPDI is negative (i.e., <0), the item has a poor fit and should be discarded.
 - If the PPDI is positive but less than 0.1, the item is not sensitive to instruction.
 - If the PPDI is between 0.1 and 0.2, the item has low sensitivity to instruction.
 - If the PPDI is between 0.2 and 1.0, the item has acceptable sensitivity to instruction.

Table 2.

Item Difficulty Index (IDI) and Pre-Post Difference Index (PPDI) Values and Interpretation of PPDI Values

Question	Item difficulty		PPDI	Interpretation of results
	Pretest	Posttest		
Question 1	0.16	0.86	0.70	Questions 1–6 have acceptable PPDI values and are very sensitive to instruction. The instruction had positive impact on participant understanding of the concepts these questions address. In
Question 2	0.22	0.82	0.60	

Question 3	0.38	0.90	0.52	general, high PPDIs indicate good instruction and/or a high-quality test item (Lan et al., 2012; Polikoff, 2010).
Question 4	0.48	0.85	0.37	
Question 5	0.48	0.79	0.31	
Question 6	0.25	0.46	0.21	
Question 7	0.70	0.81	0.11	Questions 7 and 8 have low PPDIs. It is often impossible to say whether low PPDIs are due to poor-quality test question or poor-quality instruction (Polikoff, 2010). A low PPDI also could imply that the question addresses a concept most students already had mastered before the instruction. Given that questions 7 and 8 were easy at pretest (IDI of 0.70 and 0.83, respectively) and posttest (IDI of 0.81 and 0.88, respectively), it is logical to conclude that they are addressing concepts most of the students already understood. The instructor needs to review the questions vis-à-vis the curriculum and its learning goals to evaluate whether the concept should even be covered in the instruction. Instructors should also review low PPDI questions to remove any ambiguity in wording.
Question 8	0.83	0.88	0.05	
Question 9	0.30	0.32	0.02	Question 9 is a nonsensitive question that was difficult at both pretest and posttest (IDI of 0.30 and 0.32, respectively). As earlier indicated, low PPDIs could be due to poor-quality test questions and/or poor-quality instruction (Polikoff, 2010). In general, questions are expected to be difficult at pretest and easier at posttest—after the curriculum has been enacted. However, given that the question remained difficult at posttest, it is worthwhile for the educator to review both the quality of the question and the quality of instruction. The educator should review the curriculum and its delivery to ensure that the concept addressed by the question is adequately covered and properly delivered during instruction. Additional data on instruction delivery (e.g., observation data) may help educators determine whether low PPDIs are due to poor-quality instruction (Polikoff, 2010). In this case, the educator should review the question to ensure alignment with curriculum content, and remove any ambiguity in wording.
Question 10	0.43	0.32	-0.11	Question 10 should be discarded because of its negative PPDI (Li et al., 2012).

Conclusion

As illustrated in the example presented here, the PPDI is a simple measure of pretest-posttest validity. The method for calculating and interpreting PPDIs is a reliable and robust approach to estimating the impact of a curriculum on participants. The analysis is based on frequencies and percentages and does not require any

specialized software. An educator who wishes to employ this method needs only a calculator for the analysis. The method should be a simple and useful addition to the toolboxes of Extension educators. Examining the PPDI of curriculum assessment tests would provide Extension educators with useful information for program and assessment improvement. The findings of such analyses also would help Extension educators know whether participants are gaining knowledge from a program.

References

- Arnold, M. E. (2006). Developing evaluation capacity in Extension 4-H field faculty. *American Journal of Evaluation, 27*(2), 257–269.
- Huang, Y. M., Yang, Y. H., Lin, S. J., Chen, K. C. S., Kuo, C. C., & Wu, F. L. (2015). Medication knowledge to be improved in participants in community universities in Taiwan: Outcome of a nationwide community university program. *Journal of the Formosan Medical Association, 114*(12), 1267–1279.
- Lamm, A. J., Israel G. D., & Diehl, D. (2013). A national perspective on the current evaluation activities in Extension. *Journal of Extension, 51*(1), Article 1FEA1. Available at: https://www.joe.org/joe/2013february/pdf/JOE_v51_1a1.pdf
- Lan, M. C., Li, M., Ruiz-Primo, M. A., Wang, T., Giamellaro, M., & Mason, H. (2012, April). Linking quality of instruction to instructionally sensitive assessments. Paper presented at the meeting of the American Educational Research Association. Vancouver, BC, Canada.
- Li, M., Ruiz-Primo, M. A., & Wills, K. (2012). Comparing methods to estimate the instructional sensitivity of items. Retrieved from <http://source.ucdsehd.net/deisa/4>
- Polikoff, S. (2010). Instructional sensitivity as a psychometric property of assessments. *Educational Measurement: Issues and Practice, 29*(4), 3–14.
- Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan, 89*(2), 146–150.
- Silliman, B. (2016). E-Basics: Online basic training in program evaluation. *Journal of Extension, 54*(1), Article 1TOT1. Available at: <https://www.joe.org/joe/2016february/tt1.php>
- Wise, D. K. (2017). Evaluating Extension impact on a nationwide level: Focus on program or concepts? *Journal of Extension, 55*(1), Article 1COM1. Available at: <https://www.joe.org/joe/2017february/comm1.php>
- Xie, C., Zhang, Z., Nourian, S., Pallant, A., & Bailey, S. (2014). On the instructional sensitivity of CAD logs. *International Journal of Engineering Education, 30*(4), 760–778.

Copyright © by Extension Journal, Inc. ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the [Journal Editorial Office, joe-ed@joe.org](mailto:joe-ed@joe.org).

If you have difficulties viewing or printing this page, please contact [JOE Technical Support](#)