Title	Controlling response shift bias: The use of the retrospective pre-test design in the evaluation of a master's programme
Authors(s)	Drennan, Jonathan, Hyde, Abbey
Publication date	2008-11-15
Publication information	Drennan, Jonathan, and Abbey Hyde. "Controlling Response Shift Bias: The Use of the Retrospective Pre-Test Design in the Evaluation of a Master's Programme" 33, no. 6 (November 15, 2008).
Publisher	Taylor & Francis (Routledge)
Item record/more information	http://hdl.handle.net/10197/4097
Publisher's statement	This is an electronic version of an article published in Assessment & Evaluation in Higher Education,Vol. 33, Iss. 6, 2008. Assessment & Evaluation in Higher Education is available online at: www.tandfonline.com//doi/abs/10.1080/02602930701773026
Publisher's version (DOI)	10.1080/02602930701773026

Controlling response shift bias: The use of the retrospective pretest design in the evaluation of a master's programme.

Jonathan Drennan and Abbey Hyde
University College Dublin, Dublin, Ireland

Abstract

Student self-report measures of change are widely used in evaluation research to measure the impact and outcomes of an educational programme or intervention. Traditionally the measures used to evaluate the impact of an educational programme on student outcomes and the extent to which students change is a comparison of the student's pretest scores with their posttest scores. However, this method of evaluating change may be problematic due to the confounding factor of response shift bias. Response shift bias occurs when the student's internal frame of reference of the construct being measured, for example research ability or critical thinking, changes between the pretest and the posttest due to the influence of the educational programme. To control for response shift bias the retrospective pretest method was used to evaluate the outcomes achieved from students completing a research module at master's level. The retrospective pretest method differs from the traditional pretest-posttest design in that both posttest and pretest perceptions of respondents are collected at the same time. The findings indicated that response shift bias was evident in student self-reports of change, especially in subjects the student had been previously exposed to at undergraduate level. The retrospective pretest design found that the programme had significantly greater impact on outcomes that that identified using the traditional pretest-posttest design leading to the conclusion that students may overestimate their ability at the commencement of an educational programme. The retrospective pretest design is not a replacement for the traditional pretest-posttest measures but may be a useful adjunct in the evaluation of the impact of educational programmes on student outcomes.

Introduction

Student self-report measures of change are widely used in evaluation research to measure the impact and outcomes of an educational programme or intervention. Traditionally the design used to evaluate impact is the measurement and comparison of the student's self-reported pretest scores with their posttest scores. Traditional pretest-posttest measures work on the assumption that the respondent's assessment of the measurement will not change from the pretest to the posttest. However, the respondent's perception of the construct under evaluation may change as a result of the educational intervention leading to an underreporting by the respondent of any real change occurring between pretest and posttest, this change in perception is known as response shift (Howard and Dailey 1979, Howard 1980, Goedhart & Hoogstraten 1992, Lam & Bengo 2002, Shadish *et al.* 2002). One way that has been suggested to reduce the confounding effect of this response-shift is the use of retrospective pretests when evaluating student self-reports of change. This paper reports on the use of retrospective pretest to control for response shift in the evaluation of a research module completed as part of a taught master's degree in nursing. This paper also critically evaluates the use of the retrospective pretest design and outlines the rationale for using the design in this study.

Problems with Traditional Measures of Student Change

The traditional pretest-posttest design uses the difference between the student's pretest score and their posttest score to provide a change score. In theory if the posttest score is significantly greater than the pretest score, it should indicate that change occurred on the educational variable of interest (for example problem solving, research ability, communication skills, leadership ability, critical thinking). However, traditional methods of evaluating change, such as the pretest-posttest design, may be problematic.

One major problem with the pretest-posttest design is that the student's conceptualisation or 'internal frame of reference' of the construct being measured may change (Goedhart & Hoogstraten 1992, p. 699). When using self-report pretest-posttest instruments the student may reconceptualise the construct under investigation between the pretest (time one) and the posttest (time two) (Howard 1980). This

reconceptualisation of the construct may lead the student to evaluate the construct under investigation from a different perspective at the posttest stage from the one they held at the pretest stage. This change in perspective or internal frame of reference is as a result of the student being exposed to the intervention between the pretest and the posttest leading to a shift in their response. This may result in the student using a different metric to rate themselves at time two than the one they used at time one even though measurements at time one and time two are being taken using the same instrument.

Basically, in traditional pretest-posttest designs students are required to use the same standard for measuring their ability at the beginning of a course as they are at the end of the course. Students may over-evaluate their ability or knowledge at the commencement of a programme, however following completion of the programme they may realise that their level of knowledge at the beginning of the programme was much lower than they actually estimated. This could result in there being no change in reported scores measured on a pretest scale when compared to a posttest scale. For example, a student having completed a quantitative research module at undergraduate level may estimate their knowledge of statistics as being at a level of 8 (above average) on a scale of 1 to 10 at the beginning of a research module on a master's programme. However, on completion of a research module at master's level they may realise that their knowledge of statistics following completion of their undergraduate programme was only average, however as the same scale is used at the end of the master's programme (1 to 10), they may also record 8, therefore implying that no change occurred between the commencement of the programme and the end of the programme when in fact change did occur. Therefore, student's self-report ratings of their ability at the beginning of a programme may be inaccurate (Howard & Dailey 1979). What has occurred is that students are rating their ability on a different dimension or metric at time two (posttest) than they did at time one (pretest) (Sprangers 1988). This mismatch between pretest and posttest scores is known as response shift-bias, which may result in inaccurate pretest and posttest ratings (Howard *et al.* 1979, Rohs 1999). The consequence of response shift bias is that students' pretest scores may be higher than they actually are, consequently their posttest scores may show little or no change, resulting in non-significant findings

(Umble *et al.* 2000). Therefore, the comparison of the scores from time one and time two may be misleading, inaccurate and incomparable.

The rationale underlying response shift bias is that the students' exposure to the programme leads them to a greater understanding of the construct under investigation. This in turn leads them to alter their frame of reference on the construct being measured and calls into question the internal validity of measurements taken using traditional pretest-posttest designs (Howard *et al.* 1979, Pohl 1982, Rohs 1999). Taking the example again of a student moving between a bachelor's programme and a master's programme, students may change their perceptions of their initial level of research ability between time one and time two. Following exposure to a research module of a master's programme increased understanding of the constructs to be measured would come about leading to a 'more accurate assessment of their pre-treatment levels of functioning' (Howard 1980: p. 96). The analysis of self-report outcome measures led Howard (1980; p. 100) to conclude:

> In view of the broad range of settings and instruments in which response-shifts have been observed, it seems possible that a sizable portion of the literature on program evaluation, counselling and clinical outcomes, training, group attitude, and personality research may have been influenced by response shifts.

Howard (1980) identified that respondents, after an educational intervention, self-reported little or no change in behaviour when posttest results were compared to pretests. However, these responses were not congruent with respondents' actual behaviour which in fact showed that the interventions were effective. This was evident in a communication skills workshop on dogmatism for US Air Force personnel (Howard 1980). The aim of the workshop was to decrease dogmatic tendencies in participants; however respondents' post-course measurements following the workshop showed an apparent increase in dogmatism. The rationale for this finding was that participants changed their perception of the construct of dogmatism as a result of the workshop. At the pretest stage participants tended to underestimate their dogmatic tendencies, however following the workshop the participants' perception had changed and they now rated themselves higher on dogmatism (due to a change in their conceptualisation of dogmatism) at the posttest stage even though participants, as a result of the workshop, had actually become less dogmatic.

Retrospective Pretests

To control for response shift bias it has been suggested that the retrospective pretest method (other terms used in the literature include the then-post design, thentest, or the post-then-pre design) be used in self-report measures of change (Howard *et al.* 1979, Howard 1980, Bray *et al.* 1984, Sprangers and Hoogstraten 1987, 1988a, 1988b, 1989, 1991, Sprangers 1988, 1989a, 1989b, Goedhart and Hoogstraten 1992, Umble *et al.* 2000, Rohs 2002). The retrospective pretest method differs from the traditional pretest-posttest design in that both posttest and pretest perceptions of respondents are collected at the same time. Basically the design asks the respondent to recall a point in the past and compare it to where they are now. The collection of thentest and posttest ratings at the same time leads to the reduction of response-bias due to the fact that the respondent is making the ratings at time one (thentest) and time two (posttest) from the same perspective (Howard 1980, Sprangers 1988, 1989a, 1989b). The theoretical assumption underlying the retrospective pretest method is that by asking the respondent to rate where there are now in terms of ability in relation to the construct under investigation and where they were prior to the educational intervention, they will be using the same internal frame of reference or metric to rate the construct of interest. Howard (1980) concluded that the use of retrospective pretesting could provide a more accurate indicator of respondent's change following an educational intervention than can the traditional pretest-posttest design. Objective measurements of change were found to correlate more highly with retrospective pretest designs than with pretest-posttest designs.

Retrospective pretest questioning has previously been used to evaluate both educational and social programme outcomes, these include leadership skill courses (Rohs 1999, 2002), public health education programmes (Umble *et al.* 2000, Farel *et al.* 2001) courses in statistics and research methods (Pohl 1982, Townsend *et al.* 1998, Townsend and Wilton 2003), a healthy start programme designed to prevent child abuse (Pratt *et al.* 2000), and communication skills training for medical students (Sprangers 1989a).

It was hypothesised in this study that response shift might be an issue in collecting data on the outcomes achieved as a result of a master's programme. The majority of

students undertaking a master's programme had completed either a bachelor's degree or a higher/postgraduate diploma therefore may have preconceived ideas of what study at master's level may entail. The metric on which the posttest was evaluated would change due to graduates identifying that the programme entailed more depth that previously envisaged.

Methods

Programme Evaluated

A research module of a taught masters in nursing programme was evaluated using a retrospective pretest design. The data was collected from one university, over two semesters. The content of the module included lectures on advanced quantitative and qualitative research methods with an emphasis on preparing for the development of a thesis. As well as lectures students completed workshops in statistics and the use of quantitative (SPSS) and qualitative (Nvivo) software packages. Students also had contact with a research supervisor either individually or in groups to facilitate preparation of a 20,000 word thesis. In preparation for the thesis the emphasis of teaching and supervision was on linking research theory to the practicalities of undertaking a dissertation. Therefore it was intended that the sessions would convert 'abstract conceptual knowledge into the procedural knowledge needed to conduct research and to truly understand research activity' (Murtonen & Lehtinen 2003, p. 173).

Aim of the Study

The aim of the evaluation was to measure students' self-reports of change in their ability to both understand and use research in their professional practice but also to test whether a response-shift had occurred in student's concept of research ability following exposure to a research module. Due to the fact that students had been previously exposed to research at undergraduate and higher diploma levels there was a possibility that the student's perception of the construct under evaluation (i.e. research) may change as a result of the educational intervention leading to an

underreporting by the respondent of any real change occurring between pretest and posttest.

Sample

Students from an MSc in Nursing programme in one institution were surveyed. Students surveyed had graduated between the years 2003 and 2005. A total of one hundred and twenty students were included in the study. All students responded to the pretest, with ninety-six students responding to the retrospective pretest, resulting in a response rate of eighty per cent. Students were excluded from the retrospective pretest if they had outstanding components of the master's programme to complete, therefore only those who had been awarded a masters in nursing degree were included in the follow up survey.

Instrument

The instrument was developed specifically for the master's programme and is entitled the Masters in Nursing Outcomes Evaluation Questionnaire. The section of the questionnaire reported in this paper consisted of 21-items that related to research covered in the course. Items were presented on a 7-point scale that asked participants to rate their ability from 1 indicating low ability to 7 indicating high ability. To test for response shift-bias the instrument was presented at two times and in two formats: at the beginning of the programme (time one) as pretest items only and six months after the course (time two) in the format of a posttest and a retrospective pretest. The pretest questionnaire at time one asked students to rate their ability on twenty-one aspects of research prior to commencing the programme. The posttest section of the questionnaire administered at time two asked respondents to rate where they saw themselves now as a result of completing the research component of the master's course whereas the retrospective pretest section asked the graduate to think back to the beginning of the programme and rate where they saw themselves prior to commencing the research component of the master's course. The same items appeared on both the pretest (time one) and posttest/retrospective pretest (time two) versions of the questionnaire. Respondents were therefore asked at time two to report their level of ability at present on each item following the programme (posttest) and were then asked to think back and rate themselves on each item before the programme

commenced (thentest). The rationale for adding the thentest section was to identify if response-shift bias was a confounding factor in student evaluations of change. Items for the questionnaire were developed from course documents and an extensive review of the literature that identified outcomes that should ensue following a research module at master's level. The questionnaire was tested prior to administration for face validity and content validity using the cognitive interviewing technique (Drennan 2003).

Procedure

Pretests were undertaken on the first day of the research unit. This measured student's self-reports of their current ability in a number of areas or research. Students completed the self-report posttest and the retrospective pretest six months after completing the programme by postal questionnaire. The rationale for follow-up after six months was to allow graduates time to consolidate their experience of research in their professional practice. The study was approved by the human sciences research ethics committee of the university in which the data was collected. To ensure high response rates Dillman's (2000) Tailored Design Approach was used in the postal survey component of the study. This consisted of the use of pre-letters, personalised letters, the inclusion of stamped addressed return envelopes and multiple reminder contacts.

Data Analysis

Demographic data was analysed using frequencies and measures of central tendency. Data from the pretest, posttest and retrospective pretest was analysed using a repeated measures design. Due to the relatively small sample size, ordinal level of data and non-normally distributed data (assessed by Kolmogorov-Smirnov test), Friedman's ANOVA was chosen (non-parametric test). Post-hoc testing consisted of Wilcoxon signed-rank test with *Bonferroni Correction;* 0.17 was used as the critical level of significance to prevent against the possibility of a type I error (3 comparisons .05/3 = α = .017) (Field 2005). This allowed for the comparison of pretest with posttest scores and thentest with posttest scores as well as indicating if response-shift was a factor through a comparison of conventional pretest scores with thentest scores. Effect sizes are also reported and were calculated using Pearson correlation coefficient (Field

2005, Leech *et al.* 2005). Effect sizes of $r = .10$ were considered small; of $.r = 30$ were considered medium and of .50 large (Cohen 1988).

Findings

Demographic profile of the sample.

The majority of the sample was female. The mean age was 37.9 years (*SD* 6.56). The vast majority of respondents attended their master's programme on a part-time basis. The respondents had wide experience in a variety of areas in nursing. Students held either a primary degree (mainly a Bachelor of Science in Nursing) and/or a higher/postgraduate diploma in a specialist area of nursing (for example coronary care, accident and emergency) (Table 1). All students had completed a research component as part of their undergraduate studies prior to commencing their master's degree.

Insert Table 1 About Here

Identifying Response Shift Bias

Measures of central tendency and variability for the pretest (time one - the commencement of the programme) and posttest-thentest (time two - six months following completion of the programme) are displayed in Table 2. The posttest data indicated that on all items students had positively changed in their research ability when compared to the pretest scores and thentest scores. The highest change scores were in students' ability to provide research evidence to introduce change in professional practice, ability to understand the language of research and ability to access literature relevant to their professional work. The lowest ratings related to change in ability were associated with statistical analysis, statistical problem solving and the use of statistical software packages, however statistically significant gains were also noted in these areas. Repeated measures Friedman's ANOVA identified significant differences between the mean scores on pretest, posttest and thentest data on all twenty-one items (Table 2).

Insert Table 2 About Here

To ascertain the specific differences between pretest-posttest, posttest-thentest and pretest-thentest scores and to indicate whether response shift was a factor, Wilcoxon signed rank test with Bonferroni correction was undertaken. Self-reported change was significant for both conventional pretest-posttest ratings and thentest-posttest ratings with students positively gaining in all areas of research (Table 3). However, when pretest-thentest scores were analysed it was found that students had significantly lower mean scores on fourteen items on the thentest when compared to the pretest, indicating that in these items response-shift was a factor. For example in the item 'ability to identify areas worthy of research' students rated their pretest ability at $M = 5.37$ ($SD = 1.07$) whereas on the thentest students rated their ability at only $M = 3.55$ ($SD = 1.22$) indicating that following completion of the programme students had significantly lowered their perception of their pre-programme ability. A further example of response shift was evident on the item 'ability to analyse and interpret quantitative data'; although there were significant differences between pretest and posttest scores and posttest and thentest scores, effect sizes were greater in posttest-thentest scores (effect size .43 versus .74) indicating a greater degree of change between posttest and thentest than that which occurred between pretest and posttest. Only on items that related to the use and analysis of statistics in professional practice, the ability to write findings following analysis of data, the ability to use statistical software packages and the ability to undertake research to test ideas was response shift not an issue. Furthermore, it was found that overall effect sizes were smaller for the conventional pre-test − post-test items (ranging from .24 to .81 − small to large effect, mean effect size .61) and larger for the retrospective pretest (thentest) ratings (ranging from .67 to .81 − large effects only, mean effect size .78). Mean thentest ratings were significantly lower than mean pretest ratings in fourteen items indicating that students had significantly overestimated their ability at the beginning of the programme when compared to retrospectively rating their ability at the end of the programme. This finding shows evidence of the confounding factor of response shift bias.

It is worth identifying the level of change that occurred in student's understanding and ability in research as a consequence of the research module (all comparisons will be

made between posttest and thentest ratings). Students changed substantially in all areas of research ability except in the area of statistics and in the use of qualitative software analysis packages. Although students reported statistically significant gains in these areas, the gains were less than in other areas of the programme. The lowest gains were in the students' ability to statistically analyse research data collected in professional practice, ability to use statistical and qualitative data software packages, and ability to solve statistical problems. The largest gains were in the student's ability to provide research evidence to introduce change in their professional practice, the ability to carry out a research project, ability to identify areas worthy of research, the ability to understand the language of research and the ability to critically evaluate published research.

Discussion

The rationale for the study was not only to measure the outcomes achieved as a consequence of a research module at master's level but to also ascertain whether response shift bias was an issue in measuring student self-reports of change. Therefore to control for response shift bias student change over time was measured using the retrospective pretest design. The rationale for this design was based on theories of change that identified the confounding factor of response-shift bias.

The retrospective pretest design identified that the research module evaluated had more impact on research ability than that identified using the traditional pretest – posttest design only. This finding supports Howard's (1980) contention that response shift can confound internal validity on self-report measures of change. There was evidence of response shift in a number of research areas with students significantly lowering their scores on pre-programme ability retrospectively following exposure to the programme. Although there were statistically significant differences between conventional pretest-posttest measurements, the mean difference and effect sizes were greater in the posttest-thentest (retrospective) measures. Only using the conventional pretest-posttest design would have significantly reduced the level of change self-reported by participants, thereby identifying that the educational programme may have had less impact on student change than it actually had. The findings in this study, similar to a number of studies on outcomes following education programmes,

indicated that students tended to overestimate their ability prior to the programme commencing (Hoogstraten 1982, Cantrell 2003). However, on completing the programme students recalibrated their perception and concluded that their pre-programme ability was not as high as originally thought. The theory of response shift would state that this conceptual shift occurred due to exposure to the educational programme during which students became aware of their ability and were able to accurately reconceptualise where they were at the beginning of the programme following completion of the programme. The argument underlying the use of a retrospective pretest is that that scores obtained from posttest minus thentest are more likely to accurately reflect a positive intervention effect than scores obtained from the traditional pretest-posttest method (Howard 1980, Sprangers 1988, 1989a, 1989b, Sprangers and Hoogstraten 1987, 1988a, 1988b 1989, 1991).

Although retrospective pretests are useful in identifying response shift, they are not without criticism. Howard *et al*. (1979) and Shadish *et al.* (2002) recommended that the retrospective pretests should not be used as a replacement for the conventional pretest-posttest design but should be considered as an adjunct to other methods when response shift may be an issue in self-report measures. Other problems identified with retrospective pretests include social desirability, impression management and, response bias (Lam & Bengo 2002), poor memory (Howard *et al.* 1979, Howard 1980, Lam & Bengo 2002), lack of a traditional pretest prior to the intervention (Shadish *et al*. 2002), regression to the mean (Pratt *et al.* 2000, Shadish *et al.* 2002) and maturational effects (Pratt *et al.* 2000). However, in advanced education programmes such as a master's degree it is argued that a retrospective pretest design, despite its limitations, is an effective method for measuring change in postgraduate students. This is due to the fact that students enter a postgraduate programme with preconceptions of the content of the programme based on their previous experience of exposure to constructs such as research, however, during the process of the programme students' conceptualisations change. The initial conceptualisation of the construct may have resulted in the student overestimating their ability prior to the programme commencing, which results in evidence of little or no change from the beginning of the programme to the end of the programme when traditional pretest-posttest measures are used.

The largest impacts of research on students identified using the retrospective pretest design were in relation to ability to carry out a research project, the ability to produce scholarly reports and papers, understanding of the language of research ability to develop a research instrument or questionnaire, ability to write a summary of findings from analysis of data, ability to undertake research and overall research ability. The results of this study indicated that the ability of to apply research to practice was enhanced by the programme.

The areas of lowest ability and in which response shift was not an issue were related to statistics. This finding is comparable to a wide-range of literature that has identified statistics as being particularly problematic at both undergraduate and postgraduate levels for students (Townsend *et al.* 1998, Murtonen & Lehtinen 2003). The reasons postulated for these problems include student anxiety regarding statistics (Townsend *et al.* 1998), the association of statistics with previous poor performance in mathematics during prior education (Garfield and Ahlgren 1988) and, negative attitudes towards statistics (Gal and Ginsburg 1994). Furthermore, nursing students have limited exposure to quantitative research methods and statistics at undergraduate level. Therefore response shift bias would not have been an issue in this area of research.

Conclusion

The traditional pretest-posttest method would have led to an underestimation of the impact of the research unit on student outcomes. In most cases respondents overestimated their ability, knowledge and skills in a number of areas of research prior to commencing the programme. The retrospective pretest was a more accurate indicator of change than that identified using the traditional pretest-posttest design. The use of retrospective pretest design may be justified when respondents come to an educational programme or module with some understanding of the construct, however this understanding may result in the student overestimating their ability prior to the programme commencing. The majority of students in this study had undertaken a research module at undergraduate level however their construct or metric of research changed when introduced to more advanced research areas at postgraduate level. Therefore in conclusion the retrospective pretest design is an option open to educators

in higher education who need to accurately identify the extent to which students change, especially students who have previously been exposed to the constructs being delivered.

References

Bray, J., Maxwell, S. & Howard G. (1984) Methods of analysis with response-shift bias, *Educational & Psychological Measurement,* 44, 781- 804.

Cantrell, P. (2003) Traditional vs. retrospective pretests for measuring science teaching efficacy beliefs in preservice teachers, *School Science and Mathematics*, 103, 177-185.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd edition), (New Jersey, Erlbaum).

Dillman, D. (2000) *Mail and Internet Surveys: The Tailored Design* Approach (2nd edition), (New York, John Wiley and Sons).

Drennan, J. (2003) Cognitive interviews; verbal data in the development and pretesting of questionnaires, *Journal of Advanced Nursing,* 42, 57-63.

Farel, A., Umble K. & Polhamus, B. (2001) Impact of an online analytic skills course. *Evaluation and the Health Professions,* 24, 446-459.

Field, A. (2005) *Discovering Statistics Using SPSS* (2nd Edition), (London Sage).

Gal, I. & Ginsburg, L. (1994) The role of beliefs and attitudes in learning statistics: Towards an assessment framework, *Journal of Statistics Education*, 2, 1-15.

Garfield, J. & Ahlgren, A. (1988) Difficulties in learning basic concepts in probability and statistics: implications for research, *Journal for Research in Mathematics Education,* 19, 44 – 63.

Goedhart, H. & Hoogsstraten, J. (1992) The retrospective pretest and the role of pretest information in evaluative studies, *Psychological Reports*, 70, 699-704.

Hoogstraten, J. (1982) The retrospective pretest in an educational training context, *Journal of Experimental Education*, 50, 200-204.

Howard, G., Schmeck, R. & Bray, J. (1979) Internal invalidity in studies employing self-report instruments: a suggested remedy, *Journal of Educational Measurement,* 16, 129-135.

Howard, G. (1980) Response shift bias: a problem in evaluating interventions with pre/post self-reports, *Evaluation Review*, 4, 93-106.

Howard, G. & Dailey, P. (1979) Response shift bias: a source of contamination of self-report measures, *Journal of Applied Psychology,* 64, 144-150.

Lam. T. & Bengo, P. (2002) A comparison of three retrospective self-reporting methods of measuring change in instructional practice, *American Journal of Evaluation,* 24, 65-80.

Leech, N., Barrett, K. & Morgan, G. (2005). *SPSS for Intermediate Statistics: Use and Interpretation* (2nd Edition), (New Jersey, Lawrence Erlbaum Associates).

Murtonen, M. & Lehtinen, E. (2003) Difficulties experienced by education and sociology students in quantitative methods courses, *Studies in Higher Education*, 28, 171 – 185.

Pohl, N. (1982) Using retrospectives pre-ratings to counteract response-shift confounding, *Journal of Experimental Education*, 50, 211-214.

Pratt, C., McGuigan W. & Katzev, A. (2000) Measuring program outcomes: using retrospective pretest methodology, *American Journal of Evaluation,* 21, 341-149.

Rohs, F. (1999) Response shift bias: A problem in evaluating leadership development with self-report pretest-posttest measures, *Journal of Agricultural Education*, 40, 28-37.

Rohs, F. (2002) Improving the evaluation of leadership programs: control response shift, *Journal of Leadership Education*, 1, 1-12.

Shadish, W., Cook, T. & Campbell, D. (2002) *Experimental and Quasi-experimental Designs for Generalised Causal Inference*, (Boston, Houghton Mifflin).

Sprangers, M. (1988) A further note on the necessity of including retrospective pretest in self-report pretest-posttest designs to detect training effectiveness, *Tijdschrift voor Onderwijs Research,* 13, 353-355.

Sprangers, M. (1989a) Response-shift bias in program evaluation, *Impact Assessment Bulletin,* 7, 153-166.

Sprangers, M. (1989b) Subject bias and the retrospective pretest in retrospect. *Bulletin of the Psychonomic Society*, 27, 11-14.

Sprangers, M., Hoogstraten, J. (1987) Response-style effects, response-shift bias and bogus-pipeline, *Psychological Reports,* 61, 579-585.

Sprangers, M. & Hoogstraten, J. (1988a) On delay and reassessment of retrospective ratings, *Journal of Experimental Education*, 56, 148-153.

Sprangers, M. & Hoogstraten J. (1988b) Response-style effects, response-shift bias and bogus-pipeline: A replication, *Psychological Reports,* 62, 11-16.

Sprangers, M. & Hoogstraten, J. (1989) Pretesting effects in retrospective pretest-posttest designs, *Journal of Applied Psychology*, 74, 265-272.

Sprangers, M. & Hoogstraten, J. (1991) Subject bias in three self-report measures of change, *Methodika,* 5, 1-13.

Townsend, M., Moore, D., Tuck, B. & Wilton, K. (1998) Self-concept and anxiety in university students studying social science statistics within a co-operative learning structure, *Educational Psychology*, 18, 41 – 54.

Townsend, M., Kuin Lai, M., Lavery, L., Sutherland, C. & Wilton, K. (1999) Mathematics anxiety and self concept: evaluating change using the "Then-Now" procedure, *Presentation at the Joint Conference for Research in Education,* Melbourne, December 1999.

Townsend, M. & Wilton, K. (2003) Evaluating change towards mathematics using the 'then-now' procedure in a cooperative learning programme, *British Journal of Educational Psychology,* 73, 473-487.

Umble, K., Upshaw, V., Orton, S. & Kelly, M. (2000) Using the post-then method to assess learner change, *Presentation at the American Association of Higher Education Assessment Conference*, North Carolina, June 2000.

Table 1 Demographic and Academic Profile of the Sample

Age

Mean (SD)	37.9 (6.4) years
Range	26-56 years

Years Qualified as a Nurse

Mean (SD)	16.3 (6.8) years
Range	4-36 years

Gender

Females	81 (84.4%)
Males	15 (15.6%)

Mode of Attendance

Full-time	4 (4.2%)
Part-time	89 (92.7%)
Combination of full-time and part-time	3 (3.1%)

Area of Employment

Clinical nursing	43 (44.8%)
Nurse education	36 (37.5%)
Nursing management	13 (13.5%)
Other	4 (4.1%)

Academic Qualifications*

Diploma	44 (46.3%)
Higher/Postgraduate Diploma	48 (50.5%)
Primary Degree (BSc)	70 (73.7%)
Other	13 (13.7%)

*Qualifications are prior to completing the master's degree.
Respondents may hold a number of academic qualifications.

Table 2 Pre-test, Post-test and Retrospective Pretest (thentest) Scores[1] of Research Outcomes

Item	Pretest		Retrospective Pretest				Freidman's ANOVA	
			Posttest		Thentest			
	M	SD	M	SD	M	SD	χ^2	*p*
1. Ability to carry out a research project	3.54	1.03	5.88	1.01	2.72	1.23	141.86	0.001
2. Ability to produce scholarly reports or papers	3.95	1.04	5.45	1.10	3.12	1.36	109.39	0.001
3. Ability to identify areas worthy of research	5.37	1.07	5.73	0.83	3.55	1.22	106.49	0.001
4. Understanding of the language of research	4.37	1.12	6.03	0.87	3.44	1.12	117.62	0.001
5. Ability to provide research evidence to introduce change	4.52	1.14	6.14	0.94	3.86	1.42	112.59	0.001
6. Ability to use statistics in professional practice	2.94	1.23	4.67	1.46	2.51	1.31	88.46	0.001
7. Ability to critically evaluate published research	4.36	0.86	5.82	0.96	3.51	1.27	121.95	0.001
8. Ability to develop a research instrument or questionnaire	2.66	1.24	5.31	1.45	2.67	1.44	114.29	0.001
9. Ability to analyse and interpret quantitative data	4.28	4.38	4.66	1.68	2.43	1.41	88.62	0.001
10. Ability to access literature relevant to your work	5.49	0.85	6.06	0.95	4.34	1.42	85.15	0.001
11. Ability to write a summary of findings from an analysis of data	3.42	1.23	5.57	1.10	3.26	1.35	105.99	0.001
12. Ability to statistically analyse research data collected in my professional practice	3.11	1.49	4.84	1.44	2.89	1.49	71.21	0.001
13. Ability to undertake research to test my ideas	3.42	1.38	5.43	1.23	3.18	1.53	88.67	0.001
14. Ability to publish	3.15	1.55	4.51	1.59	2.39	1.39	81.67	0.001
15. Ability to apply research to practice	5.16	1.04	5.98	0.98	4.40	1.37	72.13	0.001
16. Ability to use statistical software packages	1.65	0.97	3.82	1.91	1.77	1.21	87.05	0.001
17. Ability to use qualitative analysis software packages	2.82	1.80	1.44	0.99	1.44	0.99	45.46	0.001
18. Ability to solve statistical problems	2.64	1.36	3.82	1.77	2.06	1.33	64.77	0.001
19. Ability to judge the merit of both quantitative and qualitative approaches to research	4.42	1.39	5.76	1.12	3.40	1.53	94.15	0.001
20. Ability to analyse and interpret qualitative data	3.67	1.35	5.15	1.53	2.88	1.42	80.74	0.001
21. Overall research ability	3.72	1.06	5.57	1.06	2.78	1.12	124.81	0.001

[1]Scale scores range from 1 = low understanding/ability to 7 = high understanding/ability

Table 3 Post-hoc Wilcoxon Signed Rank Test with Effect Sizes for Differences and Response-Shift Bias Between Pretest/Posttest, Posttest/thentest and Pretest/Thentest Scores

	Item	Pre-test/Post test				Thentest/Posttest				Pretest/Thentest				
		Wilcoxon		Effect Size		Wilcoxon		Effect Size		Wilcoxon		Effect Size		Response Shift
		Z	p			Z	p			Z	p			
1	Ability to carry out a research project	7.95	0.001*	.81	L	7.89	0.001*	.81	L	4.61	0.001*	.48	M	Present
2	Ability to produce scholarly reports or papers	7.04	0.001*	.71	L	7.56	0.001*	.77	L	4.43	0.001*	.45	M	Present
3	Ability to identify areas worthy of research	2.40	0.016*	.24	S	7.90	0.001*	.81	L	7.25	0.001*	.74	L	Present
4	Understanding of the language of research	7.32	0.001*	.75	L	7.90	0.001*	.80	L	4.71	0.001*	.48	M	Present
5	Ability to provide research evidence to introduce change	7.17	0.001*	.78	L	8.13	0.001*	.82	L	3.14	0.002*	.32	M	Present
6	Ability to use statistics in professional practice	6.44	0.001*	.66	L	7.42	0.001*	.76	L	2.08	0.038^{ns}	.21	S	Not Present
7	Ability to critically evaluate published research	7.23	0.001*	.73	L	7.97	0.001*	.81	L	4.59	0.001*	.46	M	Present
8	Ability to develop a research instrument or questionnaire	7.40	0.001*	.75	L	7.65	0.001*	.78	L	0.12	0.908^{ns}	.01	S	Not Present
9	Ability to analyse and interpret quantitative data	4.29	0.001*	.43	M	7.21	0.001*	.74	L	4.88	0.001*	.50	M	Present
10	Ability to access literature relevant to your work	4.27	0.001*	.44	M	7.17	0.001*	.73	L	5.42	0.001*	.55	M	Present
11	Ability to write a summary of findings from an analysis of data	7.76	0.001*	.79	L	7.52	0.001*	.77	L	0.47	0.64^{ns}	.05	S	Not Present
12	Ability to statistically analyse research data collected in my professional practice	6.05	0.001*	.62	M	6.58	0.001*	.67	M	0.80	0.45^{ns}	.08	S	Not Present
13	Ability to undertake research to test my ideas	7.18	0.001*	.79	L	6.83	0.001*	.70	L	1.15	0.25^{ns}	.11	S	Not Present
14	Ability to publish	5.27	0.001*	.54	M	6.84	0.001*	.70	L	3.19	0.001*	.33	S	Present
15	Ability to apply research to practice	4.90	0.001*	.50	M	7.03	0.001*	.72	L	3.61	0.001*	.37	S	Present
16	Ability to use statistical software packages	6.95	0.001*	.71	L	6.33	0.001*	.65	M	0.68	0.50^{ns}	.07	S	Not Present
17	Ability to use qualitative analysis software packages	3.65	0.001*	.37	M	6.32	0.001*	.64	M	1.62	0.10^{ns}	.17	S	Not Present
18	Ability to solve statistical problems	4.72	0.001*	.48	M	6.52	0.001*	.67	M	2.85	0.001*	.29	S	Present
19	Ability to judge the merit of both quantitative and qualitative approaches to research	5.78	0.001*	.59	M	7.36	0.001*	.75	L	4.36	0.001*	.44	M	Present
20	Ability to analyse and interpret qualitative data	4.41	0.001*	.45	M	7.05	0.001*	.72	L	3.71	0.001*	.38	M	Present
21	Overall research ability	7.44	0.001*	.76	L	7.74	0.001*	.79	L	4.74	0.001*	.48	M	Present

*Bonferroni correction, significant at α = .017 level. S = Small effect size, M = Medium effect size, L = Large effect size. ns = not significant.