

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Course evaluations : a tendency to respond favourably on scales

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Emerald

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Darby, Jenny A.. 2019. "Course Evaluations : A Tendency to Respond Favourably on Scales". figshare.
<https://hdl.handle.net/2134/3300>.

This item was submitted to Loughborough's Institutional Repository by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

COURSE EVALUATIONS: A TENDENCY TO RESPOND 'FAVOURABLY' ON SCALES?

Jenny A. Darby PhD.

Department of Social Sciences, Loughborough University, Loughborough, UK. LE11 3TU

Abstract:

Purpose: A possible favourable response pattern on scaled forms used as a means of evaluating training courses is investigated. This is an important issue as scales are frequently used to collect student feedback and also to measure attitude change as a result of training courses, in universities, colleges and industry,

Methodology: In part one of the study evaluation forms were examined referring to courses were completed by 879 students attending 15 university level courses and 531 students at school. In part two a more exacting test for a favourable response tendency using attitude scales was designed. This involved 212 teachers who were asked about their willingness to include 'children with special needs' in their classes.

Findings: It was found the majority of students in part one responded at the favourable end of the evaluation scales. The same tendency was noted with the teachers in part two of the study.

Implications: Courses are likely to be evaluated favourably and therefore it is suggested that it is necessary to incorporate in any evaluation a means of comparison. Internal elements of courses need to be compared and courses should be judged in terms of how good they are when compared with other similar courses.

Key words: Students, Universities, Performance measures.

Paper Type Research Paper

Introduction:

Course evaluations are an integral part of the educational and training process. Student feedback is used by Universities, colleges, many organisations in commerce and industry involved with training and even the British Office of Standards in Education in schools. Evaluations frequently involve a number of measures but the one to be examined here is Likert style response scales. These scales typically range from words such as 'extremely good or strongly agree' through a number of check points to 'extremely bad or strongly disagree'. Researchers have noted there are a number of variables which influence student evaluations of teaching. For example Perkins, Guerin and Schleh (1990), Greenwald and Gillmore (1997) and Wachtel (1998) have noted the link between grades awarded by teachers and the evaluations of their students. Crumbley, Henry, and Kratchman (2001) noted students claimed they would mark teachers lower if they used certain less acceptable teaching techniques. Another variable examined by Defusco (1999) and Liaw and Goh (2003) examined the influence of class size and note a tendency for smaller classes to be evaluated more favourably than larger. Yet another well researched variable concerns the type of course evaluated. Bassin (1974) noted poorer evaluations tended to be given to quantitative courses and Collins (1996) reported many departmental differences were based on the academic discipline involved. Boland Liehman and Stroade (2001) and Darby (2006) have shown that students rate elective courses more favourably than required ones. McGoldrick and Schuhmann (2002) found student's personal liking for the tutor had an effect on student evaluations of that tutor. These are some of the variables which have been shown to influence student evaluations of teaching. These studies have in common the tendency to compare scores on response scales between whatever two or more variables are under investigation. They highlight variables which influence the relative favourability of evaluations but do not discuss the actual placement of responses on Likert scales. Coladarci and Kornfield (2007) reviewing the literature on student evaluations of teaching present the consensus view that the influence of variables described above is only minimal and if properly constructed and interpreted they form an important resource for improving teaching and informing personnel decisions.

The results of these student evaluation studies have in common a tendency concerning the responses on the Likert scales which is not highlighted in the reports themselves, but are observable when a closer examination of the actual data presented is carried out. For example Liaw and Goh (2003) report 'that the overall rating of all

lecturers is at least 65 per cent. and above'. Perkins, Guerin and Schleh (1990) use a seven point scale to assess teaching covering a range of characteristics such as presentation and stimulation. If the actual data reported is examined it can be seen that fourteen of the fifteen scales have mean scores at the favourable end of the scales. Again if the data presented by Coladarci and Kornfield (2007) is examined in detail all of the 27 scaled responses have a mean score at the favourable end of the five point scales.

Although student evaluation of teaching studies appear to have largely failed to draw attention to bunching at the favourable end of response scales a small number of studies into other topics have specifically mentioned this response bias. Amongst these is one by John and Robbins (1994) who asked 102 Masters in Business Administration students to rank their own performance in an interactive task and found a marked tendency to respond favourably with regard to their own performance. This study was concerned with self-esteem and is very different to a course evaluation which can be a far less personal thing. John and Robbins also looked at whether 33 of the Masters in Business Administration students considered their tutors had marked them fairly and found on a nine point rating scale a mean rating of 7.3. A total of 76 per cent of the responses were at the favourable end of the scales with this very different task. Also using Likert style scales, Schwarz et al (1991) looked at perceptions of success or failure in life. They found 80 per cent were positive responses. This, it could be argued, may reflect the positive attitudes they had towards themselves rather more than being a result of a particular response pattern on the questionnaire, but equally it could be a result of a response bias. The concern of the present study is to look specifically for evidence as to whether a favourable response bias occurs when Likert scales are used.

The idea of a possible response bias based on preference to one end of a scale could, if found to occur may rank in importance alongside other characteristics of the bias on response scales which have been well documented in the research methods texts. Wells and Marwell (1976, p. 83), for example, discuss the use of a neutral point when planning a scale. A review by Dawes and Smith (1985) and supported by Oppenheim (1992), also highlighted the advantages and disadvantages of including or omitting the neutral point. Their view is that it provides an opportunity for an evasive response style. 'Response set' is another of the factors identified by researchers including Bourque and Fielder (1995), De Vaus (1991) and Hayes (2000). This means if the scales are all presented with, for example, disagree to the left end of the scale and agree to the right respondents may complete the scale by filling in the scale points in one position on the page. Identification of these response biases provides support for the idea expressed here in the present study that the issue of response bias is one which merits investigation.

Many studies suggest that rather than being evaluated favourably there should be some unfavourable evaluations of courses. There are a number of reasons for suggesting course participants might be expected to evaluate courses unfavourably. There are individual differences in learning style (Schmidt and Moust 2000) and so students are likely to react differently to course activities; individual students' have different preferences for their lecturers as people (Vanderstoep, Fagerlin and Feenstra 2000) and preferences for presentation style differ (Grand 2000), which should all mean some students would evaluate unfavourably for some courses. Most telling is the finding by Maynard et al (2002) that students tend to blame their own failure in terms of poor course presentation. As not all students complete courses satisfactorily some unfavourable evaluations would be expected.

The aim of this study is to establish, by means of a small survey of course evaluation responses, whether there is evidence scaled responses have a favourable bias. It is not something which is mentioned in either the evaluation texts (for example Holcomb 1998, Rae 2002, and Salas et al 2003), or in the research methodology texts (for example Fowler 2002, Hayes 2000 and Shaughnessy et al 2000). It was therefore felt necessary to collect evidence to show whether this does occur.

Student evaluations typically measure feedback about the course itself and/or sometimes whether attitudes have been changed. Both these measures tend to use the scaled response format. It is intended in the present study to determine whether there is evidence scaled responses may be biased at the favourable end of scales looking at, in part one, student evaluations of teaching and in part two a response scale which questions attitude measures. Any such tendency is important because of interpretations made from response scales.

Part One

Favourable responses on a course feedback scale:

The university sample:

Evaluations from 879 undergraduates attending 15 different courses were examined. The courses selected were all taken from one university department. The university was ranked in the top twenty per cent in Britain and the

courses to be evaluated were selected on the basis of those available for evaluation during a single academic year. Table 1 provides background information about the sample which includes contributions from three levels of study, a variety of lecturers and different types of courses. The sample thus includes evaluations from a sample which includes many of the variables previously mentioned as having an impact of student evaluations.

Table 1
Showing profile of the sample included in the study

Year of study	Type of Course Quantitative or topic based Required (R) or containing student who had opted to attend as an elective (RE)	Gender of lecturer	Lecturer (L) Senior Lecturer or above (SL)	No of response forms	No of Scales scored 1 or 2
Year 1	1. Introductory topic (RE)	M	L	129	2
	2. Topic (R)	F	L	39	0
	3. Quantitative (R)	M	SL	42	1
	4. Quantitative (R)	M	L	46	0
	5. Introductory topic (RE)	M	SL	147	1
Year 2	6. Quantitative (R)	M	SL	36	0
	7. Topic (RE)	M	L	66	0
	8. Topic (R)	M	L	22	0
	9. Topic (RE)	F	SL	10	1
Year 1	10. Topic (RE)	F	L	38	0
	11. Topic (RE)	M	SL	54	0
	12. Topic (RE)	M	L	62	0
	13. Topic (RE)	F	SL	42	0
	14. Topic (RE)	F	L	50	0
	15. Topic (RE)	M	SL	93	1

The evaluation questionnaire:

This was in use in the university department and, as can be seen in table 11, consists of a series of statements to which the student is expected to respond on a five point scale from strongly agree to strongly disagree. The statements cover many aspects of the courses. These included, for example, the performance of the lecturer, the feelings of the students as to what they got out of the course and the performance of the library staff.

Results:

It can be seen from table 11 the responses do bunch towards the favourable end of the scales. The scoring system is on a five point scale with the highest score being most favourable. The second column shows the average score for those fifteen modules is higher than three. This indicates the average score on the modules is at the favourable end of the scale. The third column shows the number of scales where the average is below three for any particular module. On only eight scales is there an unsatisfactory average out of a total of 270 average scale scores over the fifteen modules. In the fourth column are details for each scale of the percentage of individual students who gave an unsatisfactory response. There are no more than 13 percent on any scale. This is the most crucial figure because it includes only those who evaluated on the negative end of the scale and excludes any who may have scored in the neutral central point on the scale. Referring back to the fourth column in table 1 where the actual courses which had a negative scale are detailed there is no indication that any particular course is less well judged than any other. A noticeable feature is that the two largest courses seem to record more criticism of the library.

Table II.
Responses from 879 evaluation forms from 15 undergraduate modules.

Statement	Mean of all modules N =15	Number of modules containing a scale for the statement which has a mean below 3	Percentage Of Individuals scoring 1 or 2 ie. below average N=879 Scoring 1 2	% fav
The module has helped me to think critically	3.79	0	0 2	98
The module has given me a good understanding of the subject	3.87	1	.3 3	96
The module has developed my interest in the subject	3.75	2	.7 6	93
The module was well organised	4.05	0	0 2	98
The way the module is delivered has encouraged me to participate	3.49	2	1 6	93
I have learned a lot from this module	3.82	1	.01 3	96
Support teaching eg tutorials/labs etc supplemented the lectures	3.72	0	1 3	96
The library has the books and resources I needed for this module	3.30	2	3 10	87
I was able to get help in the library when I needed it	3.33	0	1 4	95
The teaching rooms for this module were fit for their purpose	3.81	0	.2 3	96
Projectors, boards and screens were adequate for this module	4	0	.02 1	98
The computing facilities I needed for this module were satisfactory	3.67	0	.02 2	97
The lecturer was well prepared	4.30	0	0 1	99
The lecturer communicated clearly and effectively	4.05	0	.07 3	96
The lecturer was enthusiastic about the subject	4.18	0	.03 2	97
The lecturer used helpful teaching aids	4.05	0	.01 1	98
The lecturer was a good teacher	4.03	0	.01 3	96
The lecturer could be contacted for advice by arrangement	3.83	0	.04 2	97

The school sample:

In order to address the possibility that evaluations at a 'good university' could be expected to produce favourable evaluations school evaluations were conducted at one school for years 7, 8 and 9 with a total of 531 students. This school was classified amongst the bottom five percent. in the country for academic performance according to an Office of Standards in Education rating. This was intended to take into account the issue of favourable response patterns being a result of a good institution being judged by highly motivated students as it was unlikely these school students would be positively orientated to various aspects of their schooling due to the numerous factors, including low student motivation, which contributed to the low ranking of the school. A four

point scale was used which did not include a central or neutral point to follow the format of the Office of Standards in Education questionnaire.

Results:

78% of responses, as can be seen in table 111, occur in the two left columns with the agree responses which shows a favourable reaction to the school. The strongly negative disagree which is an unfavourable reaction to the school is rarely used. Even with this group of students at the very bottom end of the educational hierarchy they are answering positively about their school experience. The majority of the 22% unfavourable responses are on two scales, both related to student behaviour, and not directed against the school. Whilst it could be argued the students may like their school this is less likely to be the case than with the undergraduate sample as the school has such a low rating and part of the reason for this is the underachievement of the students.

Table 111
Responses of students on the 'What do you think about school?'
N=531

Please read each Statement and tick the answer that best fits what you think about your school	Strongly Agree	Agree	Disagree	Strongly Disagree	% fav
This is a good school to be at	132	366	21	10	94
I am taught well	143	366	16	5	96
Teachers expect me to work hard and do my best	173	223	10	5	74
My work is assessed helpfully so that I can see how to improve it	111	329	74	10	83
Pupils in this school behave well	10	159	276	85	32
There is an adult in this School that I can talk to If I have a problem	281	223	16	10	95
Staff treat all pupils fairly and with respect	138	270	95	26	77
I feel trusted to do things On my own	148	324	47	11	88
The school is interested In the views of its pupils	138	297	79	10	82
Worthwhile homework Is set regularly	154	281	74	16	82
There is no bullying or Racial abuse in my School	37	85	239	159	4
The school is well run	175	292	42	11	88
Total percentages for each column	26.3%	52%	16%	5.7%	

Part two

Favourable responses on an attitude scale:

Attitude measures are often used as a sign of course effectiveness. It may be seen as important that participants' beliefs are changed to match some 'corporate ideal'. The problem is that if there is a tendency to respond at the 'favourable' end of a scale then the results of a questionnaire may well reflect that response bias rather than the real views of the participants. To demonstrate this response bias it was necessary to select an attitude measure which there is some reason to believe the participants feel unfavourably towards. There are few examples of this because in cases where people generally would react negatively such as against violence there is a strong social pressure to do so. In most other cases, such as for example the need for greater efficiency in the workplace, people would be expected to respond favourably anyway.

The topic selected for this study looked at the reactions teachers expressed towards the 'inclusion' of certain types of 'children with special needs' in main stream classes. This task was chosen in the light the evidence

from the research (eg Florian and Rouse 2001 and Howe and Welner 2002) many teachers and schools are against, or at best ambivalent about including 'children with special needs' in mainstream classes. They suggest this is understandable as teachers and schools have to maintain the balance, between doing all they can to achieve the best test scores, and to provide opportunities for 'children with special needs'. Further the general tendency amongst teachers, according to Garvar-Pinhas and Schmelkin (1989), is the more contact with 'children with special needs', the more negative the attitude towards 'inclusion'. Responses, according to these researchers are therefore it is predicted likely to be negative on any response scales concerning their feelings about the inclusion of 'children with special needs' in mainstream classes. It was felt this would be an exacting test of an attitude response bias on a Likert scale. It also provided an opportunity to build into the study a more objective measure of the teacher's feelings about 'inclusion' by asking them to rank different groups of children in order of priority for 'inclusion'. This provided a measure against which the Likert scale responses could be tested. If the teachers said they were happy to have a certain type of 'child with special needs' in their classes their feelings about this could be checked against how they rank ordered their preferences. This ranking task provided an objective means of assessing each individual teacher's views about their relative willingness to accept the inclusion of various types of 'children with special needs'.

Participants:

A questionnaires were distributed by the researcher to teachers in their schools, usually in the staff room (212 responded, a 93% response rate). The teachers did not have any special involvement with 'children with special needs', other than in the course of their mainstream teaching activities. 187 reported that they had experience of teaching at least one 'child with special needs'.

The Questionnaire:

The first part of the questionnaire required the teachers to respond on Likert style scales their feelings about a number of types of 'children with special needs' being included in mainstream classes.

The format of the seven Likert style statements were as follows:

*Children who experience a **speech impairment** should be given every opportunity to be included in mainstream classes whenever possible*

Strongly agree/agree/disagree/strongly disagree

The words **speech impairment** would be substituted in the remaining six subsequent questions by one of the following: learning disability, physically impaired, behavioural problem, visual or hearing defect, a registered disability, gifted. A factor, which can influence responses, concerns the severity of the problem. Ward, Center and Bochner (1994) found the more severe the problem the less positively inclusion was regarded by teachers. In the present study the participants were instructed to consider all the seven types of special needs as 'severe'.

Each statement required a response on a four point Likert style scale. This avoided the neutral response which could well have been a popular one due to the nature of the topic. Wells and Marwell (1976 p. 83) for example argue that having a neutral response scale point provides a 'golden mean' or the opportunity for an evasive response style. The terms on the Likert scale were presented alternately with agree on the left and then on the right to avoid 'response set'. In addition the order of presentation of the problems was varied by printing many versions of the questionnaire with the statements re ordered. This was to ensure there was no 'order effect' which could imply any particular problem was more important.

The second part of the questionnaire required the teachers to rank order the above categories in priority order for inclusion and the second ranking task required them to rank the problems from one to seven in order of the need to allocate resources. The order of presentation of the types of 'children with special needs' was varied randomly on the questionnaires. The ranking approach makes adopting socially acceptable responses less easy, and forces the participants to indicate their preferences. The ranking task was used because, according to John and Robbins (1994), a ranking task provides 'an explicit context of comparison' and 'also eliminates any potential differences in scale usage' so that judgements can be made on a specified base line. Any link between the scores on the Likert scales and the rank ordering of the 'children with special needs' could be observed

Results:

In addition to the examination of the profile of Likert responses the scores on the Likert response statements were correlated with the scores on the ranking task. This was to determine whether responses on the Likert scales represented the views expressed on the ranking task. This acted as a control measure which could not be subject to a favourable pattern of responses.

Responses to the statement; *Children who experience a should be given every opportunity to be included in mainstream classes whenever possible*

As can be seen in table IV reading from left to right, in the second column the Likert style scores show that none of the categories were scored below the 2.5 mid point level. In column three on the Likert scales by far the greatest number of responses are on the agree end of the scale as was predicted. Only 12% of the answers were on the negative end of the Likert scales. This means that 88% of the responses indicate a favourable reaction to the idea of 'inclusion' for the various types of problem. When, however, the ranking task is considered it is clear in the fourth column of table IV some of the categories are ranked very much lower than others. The ranking task suggests some 'children with special needs' are consistently seen as being more acceptable to be included than others. This is not made clear by the Likert scores. Column five shows the correlations between individual scores on the two tasks are all very low. This might be indicating that individuals are not responding consistently on the two scales. They tend to be scoring favourably on the Likert scales to all 'children with special needs' whereas on the ranking task it is clear they do not have a similar view. This difference really does highlight the problem of favourable bias with response scales. They are typically used in course evaluations to compare between variables such as class size, type of course type of tutor etc. The interpretation made from these studies tends to be that there are differences but rarely that courses are unsatisfactory. Here again in a very different context the results of this questionnaire about inclusion show the teachers are favourably disposed towards all types of children which is particularly surprising. The literature indicates this would not be expected. It also suggests teachers with experience of having 'children with special needs' in their classes are unlikely to be favourably disposed towards their inclusion in future. In the sample included in this study 88% of the teachers had some previous experience of those children in their classes. The pattern of response on the Likert scales bears little or no relationship to their willingness to accept certain types of children when forced to rank them. If this is transposed to the course evaluation situation the implications are serious.

Scales sores allow students or in this instance the teachers to respond favourably to all factors, some more than others but they are not compelled to adopt a negative strategy. When forced to rank they are well able to do so and place some factors at the bottom of the list. In other words the use of scales allows the respondent to avoid being negative. This raises the point that if scaled responses fail to produce negative reactions, although their use for comparison purposes is acceptable, their use as a measure of approval by students of courses is suspect. It is not being argued here that response scales should not be used merely that one of the limitations is that responses tend to be at the favourable end of scales. This method of evaluation may not really be suitable as a means of identifying unsatisfactory courses.

Table IV
Participants view of the problems considered most suitable for inclusion.

Type of 'special need' N= 212	Likert Score Mean, Standard deviation (in brackets) High score positive response	Number scoring each point scale. Score of 1 or indicates a negative response 1 2 3 4	Ranking score. Mean, Standard deviation (in brackets) Low score high ranking	Correlation between Likert scale and ranking score *
Other (asthma, arthritis, epilepsy, diabetes, haemophilia etc)	3.83 (0.38)	0 0 36 176	1.78 (1.26)	-.078
Physical	3.55 (0.54)	0 4 88 120	3.07 (1.70)	-.258
Speech	3.52 (0.55)	0 6 90 116	3.19 (1.50)	-.234
Gifted	3.38 (0.59)	0 12 108 92	3.58 (1.36)	-.212
Hearing/visual impair	3.26 (0.38)	0 12 132 68	5.31 (1.32)	-.074
Learning impairment	3.22 (0.69)	0 24 124 64	5.53 (1.47)	-.385
Behavioural	2.72 (0.79)	12 68 100 32	5.53 (1.57)	-.262

*The correlation is expressed as a minus, that is an inverse correlation simply because low score on ranking task is a preferred choice and a high score on the Likert scale reflects a positive response.

Discussion

The results of the undergraduate and school evaluations show the students respond favourably to the courses in which they have taken part. From the evidence of their responses to the evaluations, on virtually all the measures, scores are above the mid point on scales used. It must be comforting for the organisers of these courses to have noted the responses are this favourable! An alternative explanation is that the students are merely responding with a favourable bias on the Likert scales. The 'good evaluations' may merely reflect this

tendency. This means that feedback about individual courses is more likely to be favourable than unfavourable. Judging any individual course on the basis of feedback evaluations would not therefore appear to be a useful exercise. There are two ways in which any 'favourable response bias' can be taken into account. The first is to compare scores on evaluation forms from one course with those of another similar course. This would provide the opportunity for courses to be ranked in terms of participants' reactions. The second is to look at evaluations of different elements within a particular course, such as method of presentation or use of case studies, and compare them. That way the relative strengths or weaknesses of these elements within a particular course can be judged.

Part two of the study which involved the teacher questionnaires shows how responses on attitude scales also tend to lean towards the favourable end. Again it is suggested here this reflects a 'response bias'. The teachers answered positively on the Likert scales indicating they feel 'children with special needs' are suitable for 'inclusion' into mainstream classes even though the research evidence and the ranking task suggests this is too simplistic an interpretation. Respondents are choosing the favourable half of the scale on this task, just as the students in the university and school samples did on the various course evaluation scales.

The findings of the present study provide support for the view observed in very different contexts by John and Robbins (1994) and Schwarz et al (1991) that responses on Likert style scales tend to be placed at the favourable end. Favourable evaluations of courses do not necessarily mean the courses are good. These results suggest it is as likely to be a consequence of the way in which people respond on Likert scales. It could be argued it would be a far better judgement on a course, if the student were to be asked to rank that course in comparison to others. This would mean that rather than being allowed to award an above average score or response to each course, they would say how good the course was as compared to others. The difficulty in doing this is some topics are more popular than others (Bassin 1974) and so the selection of courses for comparison purposes needs to take this into account.

The tendency to respond at the positive end of a Likert scale should rank in importance along side other characteristics such as 'mid point and 'response set'. Evaluations with scores above the mid point on scales may be more of a feature of the manner in which individuals typically tend to respond on scales than to the quality of the course. This is important because of the emphasis given to student evaluations of teaching.

Acknowledgement:

This empirical study is integrated in the author's own doctoral thesis.

References:

- Bassin, W. M. (1974) A note on the biases in students' evaluations of instructors'. *Journal of Experimental Education* 43: 16-17
- Boland, M. Lehman, E. and Stroade, J. (2001) A comparison of curriculum baccalaureate degree programs in agribusiness. *The International food and agribusiness management review*. 225-235.
- Bourque, L.B. & Fielder, E.P. (1995). *How to conduct self administered and mail surveys*. California: Sage Publications.
- Coladarci, T. and Kornfield, I. (2007) Rate My Professors.com versus formal in-class student evaluations of teaching. *Practical Assessment Research and Evaluation*. 12. 6: 2-15
- Collins, J.W. (1996) intellectual motivation and its relationship to selected characteristics of collegiate business and liberal arts majors. *Dissertation Abstracts International Section A. Humanities and Social Sciences*. 56. (10-A) 3886.
- Crumbley, L. Henry, B.K. . and Kratchman, S.H. (2001) Students; perceptions of the evaluation of college teaching. *Quality Assurance in Education*. 9.4: 197-207.
- Darby, J.A. (2006) The effects of the elective or required status of courses on student evaluations. *Journal of Vocational Education and Training*. 58. 1: 19-29.
- Dawes, R.M. and Smith, T. (1985) Attitude and opinion measurement: In Lindsay, G. and Aronson, E. (Eds) *Handbook of Social Psychology*. 509-566. New York: Random house.

- Defusco, M.B. (1999) An exploration of the relationship between class size and students' ratings of teaching quality at the University of Phoenix: Do adult students make a difference? *Dissertation Abstracts International section A: Humanities and Social Sciences*. 60. 6-A. 1901.
- De Vaus, D.A. (1991) *Surveys in social research*. (3rd Edition) London: Unwin Hyman.
- Florian, L. and Rouse, M. (2001) Inclusive practice in English Secondary schools: lessons learned *Cambridge Journal of Education* 31. 3: 399-412.
- Fowler, F.J. (2002) *Survey research methods*. London: Sage.
- Garvar-Pinhas, A. and Schmelkin, L.P. (1989) Administrators and teachers' attitudes towards mainstreaming. *Remedial and Special Education*. 10: 38-43
- Grand, L.C. (2000) *The Workplace Skills: Presentation guide*. New York: John Wiley and Sons Inc.
- Greenwald, A.G. and Gillmore, G.M. (1997) Grading leniency is a removable contamination of student ratings. *American Psychologist*, 52: 1209-1217.
- Hayes, N. (2000) *Doing psychological research*. Buckingham. Philadelphia: Open University Press
- Holcomb, J. (1998) *Training evaluation made easy: Making your training worth every penny*. London: Kogan Page.
- Howe, K.R. and Welner, K.G. (2002) School choice and the pressure to perform: Déjà vu for children with Disabilities? *Remedial and Special Education*. 23.4: 212-221
- John, O.P. and Robbins, R.W. (1994) Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*. 66. 1: 206-219.
- Liaw, S.H. and Goh, K.I. (2003) Evidence and control biases in student evaluations of teaching. *International Journal of Educational Management*. 17. 1: 37-43.
- McGoldrick, K. and Schuhmann, P.W. (2002) Instructor gender and student registration: An analysis of preferences. *Education Economics*. 10. 3: 241-260.
- Maynard, D.C., Bachiochi, P.D. and Luna, A.C. (2002) An evaluation of industrial/organizational psychology teaching modules for use in introductory psychology. *Teaching of Psychology* 29.1:39-43
- Oppenheim, A.N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*. New York: Pinter Publications.
- Perkins, D. Guerin, D. and Schleh, J. (1990) Effects of grading standards information, assigned grade, and grade discrepancies on students' evaluations. *Psychological Reports*. 66: 635-642.
- Rae, L. (2002) *Assessing the value of your training: The evaluation process from training needs to the report to the board*. Aldershot, UK: Gower Publishing Co.
- Salas, E. Milham, L.M. and Bowers, C.A. (2003) Training evaluation in the military: Misconceptions, opportunities, and challenges *Military Psychology* 15.1: 3-16
- Schmidt, H.G. and Moust, J.H.C. (2000) Factors affecting small-group tutorial learning: A review of research. In Evensen, D.H. and Hmelo, E. (Eds) *Problem-based learning: A research perspective on learning interactions*. 19-51. Mahwah, N.J. US: Lawrence Erlbaum Assocs. Publishers.
- Schwarz, N. Knauper, B, Hippler, H.J. Noelle-Neumann, E and Clark, F.(1991) Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*. 55: 570-582.
- Shaughnessy, J.J. Zechmeister, E.B. and Zechmeister, J.S. (2000) *Research methods in psychology*. (5th edition). Boston: McGraw-Hill

VanderStoep, S.W. Fagerlin, A. and Feenstra, J.S. (2000) What do students remember from Introductory Psychology? *Teaching of Psychology*. 27.2: 89-92

Ward, J. Center, Y. and Bochner, S. (1994) A question of attitudes: integrating children with disabilities into regular classrooms? *British Journal of Special Education*. 21. 1: 34-39.

Wachtel, K.H. (1998) Student evaluation of college teaching effectiveness: a brief review. *Assessment and Evaluation in Higher education*. 23. 2. 191-211.

Wells, L.E. and Marwell, G. (1976) *Self esteem: Its conceptualisation and measurement*. Beverley Hills: Sage Publications.

Corresponding author

Jenny A Darby can be contacted at J.A.Darby@lboro.ac.uk

Biographical note:

Jenny Darby has a varied career in Education. She taught in an inner city Comprehensive school for 25 years completing her service as Head of Science. For two years during a secondment she was responsible for a county wide training programme for head teachers and senior teaching staff. She has recently been running training workshops in teaching skills for postgraduates at Loughborough University, where she completed her doctorate. Her main research interest concerns evaluation of training programmes.

E mail J.A.Darby@lboro.ac.uk