

# **Changes in the distribution of response to the 2010 CEQ resulting from changes to response scale labels**

Graduate Careers Australia

June, 2011

## Background

From its introduction in the early 1990s through to 2009, respondents to the Course Experience Questionnaire (CEQ) were asked to express the extent to which they agree or disagree with a range of Likert-type statements concerning their higher education student experience, measured on a five-point response scale with only the ‘strongly disagree’ and ‘strongly agree’ endpoints labelled, as it was originally developed. Over time, some concerns from within the Australian higher education sector were raised with regard to the possibility that graduates were misreading the direction of the response scale and selecting ‘strongly disagree’ when in fact they meant to select ‘strongly agree’ and vice versa. Positive comments in the ‘best aspects’ and ‘needs improvement’ fields accompanying negative CEQ scores (and vice versa) were provided as evidence of this phenomenon. In response to these concerns, the Survey Reference Group (SRG)<sup>1</sup> at its July 2009 meeting resolved to label all five points of the CEQ response scale.<sup>2</sup> To this end, the 2010 CEQ was administered with the response scale points labelled ‘strongly disagree’, ‘disagree’, ‘neither agree nor disagree’, ‘agree’ and ‘strongly agree’.<sup>3</sup> When responses to the 2010 CEQ were examined, an unusual pattern emerged: the scale and item scores for the 2010 CEQ were found to be higher than those in previous years and represented a clear departure from the CEQ time series (Figure 1). The change to the CEQ response scale was immediately flagged as a potential cause.

---

<sup>1</sup> The SRG is an advisory committee reporting to the GCA Board.

<sup>2</sup> Another proposal would have seen the endpoint labels changed to ‘disagree strongly’ and ‘agree strongly’.

<sup>3</sup> Prior to this, the five points of the response scale were interpreted as, but not labelled as, ‘strongly disagree’, ‘disagree’, ‘undecided’, ‘agree’ and ‘strongly agree’ (GCA and ACER, 2010).

The differences between response scales with endpoint-only labelling and full labelling are well addressed in the literature.<sup>4</sup> From a scale design perspective, endpoint-only labels seem more in line with an interval scale assumption (Weijters, Cabooter and Schillewaert, 2010), although labelling all points on a response scale is typically associated with higher reliability (e.g., Krosnick and Berent 1991; Weng, 2004). Previous research (e.g., Wildt and Mazis, 1978) has shown that respondents use the labels provided to them when they are mapping judgements to response scales, so labels denoting agreement and disagreement make the valence of a negative/positive response more explicit (Weijters, Cabooter and Schillewaert, 2010). As survey respondents generally have a desire to be agreeable (e.g., McClendon, 1991), labelling all points on the response scale is likely to shift the distribution of responses to the positive side.<sup>5</sup> Moreover, the greater accessibility of labelled response categories is also likely to cause a shift in the distribution of responses to the middle categories relative to a scale with only the endpoints labelled (Simonson, 1989).

This discussion paper has been prepared by GCA to investigate the nature of this change in response to the CEQ. Because this response scale labelling affected all respondents to the 2010 CEQ, we are forced to use CEQ responses from earlier survey rounds as a basis for comparison. This approach is less than ideal,<sup>6</sup> as it is practically impossible to distinguish between changes due to the response scale labelling and changes due to genuine improvements in the student experience at a particular institution. As shown in Figure 1, however, yearly changes in CEQ scores are typically incremental and, as a result, it is fairly safe to assume that any large changes in CEQ scores between 2009 and 2010 are a result of scale labelling.<sup>7</sup> Moreover, the composition of the sample remained fairly consistent between 2008 and 2010 (see Table A at the end of this paper), which suggests that changes in the nature of response are not due to changes in the composition of the sample.

---

<sup>4</sup> Weijters, Cabooter and Schillewaert (2010) ably survey the literature.

<sup>5</sup> Guterbock and Hubbard (2000) found that fully labelling a five-point satisfaction scale caused an upward shift in the response distribution relative to a scale with only the endpoints labelled.

<sup>6</sup> Ideally, a small-scale randomised trial should have been undertaken prior to the change being made; however, the change was seen as immediately necessary and followed long-term advocacy from the sector for labelling.

<sup>7</sup> In the interests of comparability, our analysis is based solely on CEQ responses gathered by paper or online survey. Responses collected by telephone in 2010 have been excluded to eliminate this as a possible cause. The issue of telephone collection of CEQ responses will be addressed in a separate paper later in 2011.

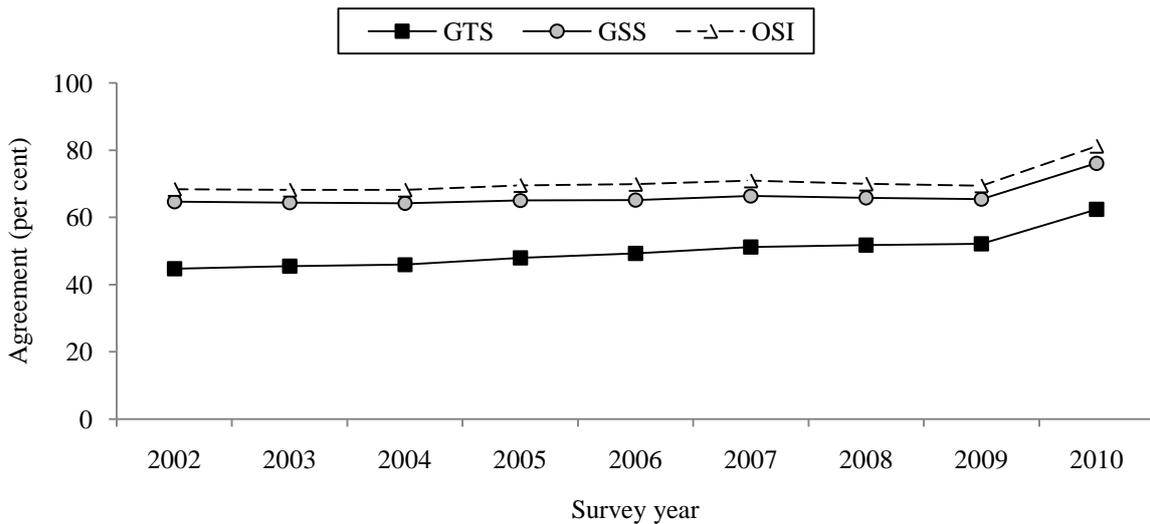


Figure 1. Trends in CEQ scale mean percentage agreement (bachelor graduates), 2002-10

#### Distribution of CEQ item responses

We begin by examining the distribution of responses across the five response categories for each of the core CEQ items in 2008, 2009 and 2010. This is presented in Table 1. Recall that in 2008 and 2009 only the endpoints were labelled, while in 2010 all five points were labelled. A number of changes in the distribution of responses are immediately evident. Firstly, respondents were much more likely to select the point labelled ‘agree’ in 2010 than they were to select the unlabelled fourth point (relative to ‘strongly disagree’) in 2009. This same pattern was observed for all 13 core CEQ items, with the largest changes on GTS01 and GTS03 (14.8 pp and 14.4 pp respectively).<sup>8</sup> The smallest such change between 2009 and 2010 was still 8.1 pp (GTS27), which dwarfs the *largest* change of 1.3 pp observed between 2008 and 2009 (GSS23). A shift away from the point labelled ‘disagree’ was also observed, although this was not as extreme. This general shift towards the positive side of the response distribution is consistent with the literature concerning respondents’ desire to be agreeable when they are presented with labelled response categories.

<sup>8</sup> The abbreviation ‘pp’ denotes percentage points.

Table 1. *CEQ item response percentages (bachelor graduates), 2008-10*

Item	2008					2009					2010				
	SD	--	--	--	SA	SD	--	--	--	SA	SD	D	N	A	SA
GTS01	3.6	14.4	31.2	37.4	13.4	3.7	14.2	30.9	36.8	14.4	2.4	11.0	22.3	51.6	12.8
GTS03	3.4	14.4	27.4	40.7	14.2	3.6	14.2	27.3	39.9	15.0	2.1	10.3	19.3	54.3	14.0
GTS10	3.9	13.2	31.1	35.9	16.0	3.9	13.3	30.5	35.8	16.6	2.4	9.7	26.8	45.1	16.1
GTS15	3.2	11.2	32.3	38.9	14.3	3.3	11.3	31.8	38.7	14.9	2.2	8.0	27.4	47.8	14.6
GTS16	3.4	11.7	30.4	38.7	15.7	3.5	11.6	29.9	38.6	16.3	2.2	8.4	25.7	47.7	15.9
GTS27	5.2	15.5	34.1	32.2	13.0	5.4	15.5	33.5	31.8	13.8	3.7	12.7	31.3	39.9	12.3
GSS06	4.4	13.1	25.1	38.9	18.5	4.4	12.9	25.9	37.9	18.9	2.4	9.2	20.7	49.7	18.1
GSS14	2.8	7.4	19.5	42.0	28.3	3.0	7.7	19.3	41.1	28.9	1.2	3.9	14.0	52.7	28.2
GSS23	2.6	8.1	22.7	44.7	22.0	2.6	8.4	22.7	43.4	22.9	1.2	4.3	16.5	56.0	22.0
GSS32	3.7	8.4	18.0	39.0	30.9	3.9	8.4	18.2	38.3	31.2	1.9	5.3	13.6	49.3	30.0
GSS42	2.6	8.5	26.4	42.7	19.7	2.8	9.0	26.1	42.2	19.9	1.4	5.2	22.3	52.4	18.7
GSS43	2.7	7.7	21.3	44.7	23.7	2.9	8.0	21.3	43.6	24.2	1.4	4.4	16.4	55.4	22.4
OSI49	3.4	8.1	18.5	46.3	23.6	3.6	8.3	18.7	45.1	24.3	2.0	5.0	12.2	57.6	23.1

Notes. GTS01 = 'The staff put a lot of time into commenting on my work'; GTS03 = 'The teaching staff normally gave me helpful feedback on how I was going'; GTS10 = 'The teaching staff of this course motivated me to do my best work'; GTS15 = 'My lecturers were extremely good at explaining things'; GTS16 = 'The teaching staff worked hard to make their subjects interesting'; GTS27 = 'The staff made a real effort to understand difficulties I might be having with my work'; GSS06 = 'The course helped me develop my ability to work as a team member'; GSS14 = 'The course sharpened my analytic skills'; GSS23 = 'The course developed my problem-solving skills'; GSS32 = 'The course improved my skills in written communication'; GSS42 = 'As a result of my course, I feel confident about tackling unfamiliar problems'; GSS43 = 'My course helped me to develop the ability to plan my own work'; OSI49 = 'Overall, I was satisfied with the quality of this course'. Responses provided by telephone excluded.

Secondly, respondents were much less likely to select ‘neither agree nor disagree’ in 2010 than they were to select the unlabelled midpoint in 2009. This was observed for all 13 core CEQ items, with the largest changes occurring for GTS01 and GTS03 (8.6 pp and 8.0 points pp respectively). By contrast, the *largest* such change observed between 2008 and 2009 was just 1.8 percentage points (GSS43). Curiously, the midpoint of the CEQ response scale has conventionally be interpreted (but not labelled) as ‘undecided’; however it was subsequently labelled as ‘neither agree nor disagree’. It is possible that the change in response to this category between 2009 and 2010 could be attributable to a change in its interpretation following the addition of labelling, or could reflect the ambiguity of the response scale midpoint when unlabelled. Finally, it should be noted that the fully-labelled response scale elicited more consistent responses, as shown by the item descriptive statistics in Table 2.

Table 2. *CEQ item descriptive statistics (bachelor graduates), 2008-10*

Item	2008			2009			2010		
	N	Mean	St. Dev.	N	Mean	St. Dev.	N	Mean	St. Dev.
GTS01	71,957	21.3	50.5	73,676	22.0	51.0	75,271	30.7	46.2
GTS03	71,944	24.0	50.6	73,653	24.3	51.2	75,309	33.9	45.6
GTS10	71,785	23.5	51.6	73,548	23.9	52.0	75,234	31.4	47.2
GTS15	71,786	24.9	48.9	73,515	25.2	49.3	75,265	32.3	45.1
GTS16	71,828	25.9	50.1	73,540	26.3	50.4	75,236	33.3	45.9
GTS27	71,641	16.2	52.5	73,400	16.6	53.2	75,175	22.1	49.3
GSS06	71,852	26.9	53.5	73,593	27.1	53.6	75,236	36.0	47.2
GSS14	71,881	42.7	50.2	73,572	42.6	51.0	75,261	51.4	41.4
GSS23	71,786	37.8	48.5	73,458	37.8	49.2	75,210	46.6	40.7
GSS32	71,783	42.5	53.2	73,447	42.3	53.7	75,213	50.2	45.1
GSS42	71,757	34.2	48.5	73,456	33.7	49.2	75,209	40.8	42.2
GSS43	71,752	39.5	48.8	73,471	39.1	49.7	75,191	46.5	41.4
OSI49	71,723	39.3	50.2	73,425	39.2	50.9	75,158	47.4	42.8

*Note.* Statistics based on a five-point response scale coded as -100, 50, 0, 50 and 100 as per the conventional CEQ reporting metric. Responses provided by telephone excluded. Item labels as per Table 1 footnote.

From the response distributions presented in Table 1, it appears that the most notable change in the nature of response to the 2010 CEQ was an upward shift from the midpoint (‘neither agree nor disagree’) to the fourth point towards the agreement end of the response scale (‘agree’).

## Conclusions

It is clear from the evidence that a shift in the distribution of responses occurred between the 2009 and 2010 CEQ rounds and that, all else being approximately equal,<sup>9</sup> this change was likely due to the addition of labels to the intermediate points of the CEQ response scale. In line with the literature, fully labelling the response scale caused an upward shift in the distribution of responses to the CEQ items. An immediate consequence of this change is that 2010 essentially represents the beginning of a new CEQ time series, as the results obtained using the fully-labelled response scale are not comparable with those obtained using the conventional CEQ response scale. The inconvenience of a break in time series notwithstanding, we believe that this change should be viewed as a positive development in terms of the CEQ. Firstly, there is evidence that the full-labelled response scale elicits more consistent responses and, secondly, fully labelling the CEQ response scale brings the instrument more in line with the Postgraduate Research Experience Questionnaire (PREQ), which has employed a fully-labelled five-point Likert-type response scale since its introduction in 1999.<sup>10</sup>

It should be noted that, as part of its *Advancing Quality in Higher Education* initiative,<sup>11</sup> the Department of Education, Employment and Workplace Relations (DEEWR) has indicated that the CEQ will be reviewed as part of a major overhaul of the Australian Graduate Survey (AGS) that will take place over the coming months. A reassessment of the CEQ response scale, among other aspects of the instrument, will likely form part of this review process.

---

<sup>9</sup> We checked the robustness of our findings by estimating a series of logistic regression models using data from the 2008, 2009 and 2010 CEQ rounds. We controlled for broad field of education and institution and used the 2008 round as a basis for comparison. We tested whether graduates in 2010 (1) were more likely to provide a positive response and (2) were less likely to respond at the midpoint. The results of this logistic regression analysis supports our previous findings that graduates in 2010 were more likely to respond positively and less likely to respond at the midpoint than graduates in 2008 and, notably, that these same response patterns did not change markedly between 2008 and 2009. These regression results are not presented in this paper in the interest of brevity but are available from GCA on request.

<sup>10</sup> The PREQ response scale includes an additional 'does not apply' category.

<sup>11</sup> See DEEWR (2011).

## References

- DEEWR. (2011). *Advancing Quality in Higher Education Information Sheet*. Canberra: Department of Education, Employment and Workplace Relations.
- GCA & ACER. (2010). *Graduate Course Experience 2010*. Melbourne: Graduate Careers Australia.
- Guterbock, T.M., and Hubbard, R.A. (2000). *Anchors away: The effect of verbal anchors on responses to a satisfaction scale*. Paper presented at the American Association for Public Opinion Research Annual Conference, Portland, OR.
- Krosnick, J.A., and Berent, M.K. (1991). *Memory for political information: The impact of attitude importance on selective exposure, selective elaboration, and selective recall*. Paper presented at the Society for Experimental Social Psychology Annual Meeting, Columbus, OH.
- McClendon, M.J. (1991). "Acquiescence and recency response-order effects in interview surveys", *Sociological Methods and Research*, 20(1), 60–103.
- Simonson, I. (1989). "Choice based on reasons: The case of attraction and compromise effect", *Journal of Consumer Research*, 16(2), 158–174.
- Weijters, B., Cabooter, E., and Schillewaert, N. (2010). "The effect of rating scale format on response styles: The number of response categories and response category labels", Vlerick Leuven Gent, Working Paper No. 2010/07.
- Weng, L.J. (2004). "Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability", *Educational and Psychological Measurement*, 64(6), 956–972.
- Wildt, A.R., and Mazis, M.B. (1978). "Determinants of scale response: Label versus position", *Journal of Marketing Research*, 15(2), 261–267.

Table A. CEQ response characteristics (bachelor graduates), 2008-10

		2007		2008		2009	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Broad Field of Education	Natural and Physical Sciences	7,412	10.3	7,978	10.9	7,275	9.6
	Information Technology	2,965	4.1	2,538	3.5	2,296	3.0
	Engineering and Related	3,829	5.3	3,915	5.3	4,376	5.8
	Architecture and Building	1,469	2.0	1,350	1.8	1,696	2.2
	Ag., Environ. and Related	1,192	1.7	1,075	1.5	1,172	1.6
	Health	9,516	13.2	10,259	14.0	11,141	14.8
	Education	5,289	7.4	5,268	7.2	5,453	7.2
	Management and Commerce	17,575	24.4	17,618	24.0	18,961	25.1
	Society and Culture	17,038	23.7	17,284	23.5	16,951	22.5
	Creative Arts	5,602	7.8	6,116	8.3	6,096	8.1
	Food, Hospitality and Pers. Serv.	25	0.0	10	0.0	14	0.0
	Mixed Field Programmes	9	0.0	5	0.0	19	0.0
Means of financing study	HECS paid upfront	12,521	17.5	12,157	16.5	10,542	14.0
	HECS deferred some or all	45,437	63.6	47,026	63.8	48,656	64.8
	International fee-paying student	10,249	14.3	10,894	14.8	11,646	15.5
	Australian fee-paying student	3,218	4.5	3,630	4.9	4,156	5.5
	APA or RTS research student	54	0.1	15	0.0	35	0.0
Attendance	Mainly full time	60,808	85.0	62,572	84.8	65,424	86.9
	Mainly part time	10,728	15.0	11,253	15.2	9,879	13.1
Study mode	Internal	61,233	85.6	62,987	85.3	65,367	86.8
	External	4,567	6.4	4,620	6.3	4,583	6.1
	Mixed mode	5,763	8.1	6,249	8.5	5,347	7.1
Sex	Male	27,246	37.7	28,309	38.3	28,592	37.9
	Female	44,837	62.1	45,564	61.6	46,764	62.0
	Unknown	110	0.2	78	0.1	102	0.1
Age group	Under 25	49,039	68.1	50,475	68.3	51,994	69.1
	25 to 29	11,267	15.6	11,822	16.0	11,853	15.7
	30 to 39	6,352	8.8	6,187	8.4	6,226	8.3
	40 to 54	4,603	6.4	4,500	6.1	4,426	5.9
	55 and over	787	1.1	886	1.2	788	1.0
ATSI	Not ATSI	71,513	99.3	73,276	99.4	73,264	99.3
	ATSI	484	0.7	454	0.6	492	0.7
Residency	Australia	62,442	86.8	63,526	86.1	64,262	85.2
	International	9,528	13.2	10,219	13.9	11,132	14.8
NESB	English spoken at home	56,380	78.9	57,789	78.4	58,351	77.7
	Other language spoken at home	15,104	21.1	15,928	21.6	16,780	22.3
Disability identification	Yes	1,804	2.5	1,944	2.6	1,915	2.5
	No	69,646	97.5	71,708	97.4	73,240	97.5
Final year paid work	Yes	55,046	78.1	56,502	77.7	55,016	74.0
	No	15,443	21.9	16,181	22.3	19,333	26.0
Paid work status	In full-time work	39,046	55.1	35,570	48.8	34,125	46.0
	In part-time work	19,026	26.9	22,042	30.2	23,288	31.4
	Not working	12,742	18.0	15,329	21.0	16,831	22.7

Note. Responses provided by telephone excluded. Subsections may not add to 100.0% due to rounding.