Total Survey Error & Institutional Research: A case study of the University Experience Survey

Sonia Whiteley

Paper presented at the 2014 Australasian Association for Institutional Research (AAIR) Forum, Melbourne, Australia

Abstract

Total Survey Error (TSE) is a component of Total Survey Quality (TSQ) that supports the assessment of the extent to which a survey is 'fit-for-purpose'. While TSQ looks at a number of dimensions, such as relevance, credibility and accessibility, TSE is has a more operational focus on accuracy and minimising errors. Mitigating survey error involves finding a balance between a achieving a survey with minimal error and a survey that is affordable. It is also often the case that addressing one source of error can inadvertently increase another source of error.

TSE provides a conceptual framework for evaluating the design of the University Experience Survey (UES) and offers a structured approach to making decisions about changing and enhancing the UES to support continuous improvement. The implications of TSE for institutional research will be discussed using the UES as a case study.

Total Survey Error

> "Total Survey Error refers to the accumulation of all errors that may arise in the design, collection, processing, and analysis of survey data. A survey error is defined as the deviation of a survey response from its underlying true value." (Biemer, 2010)

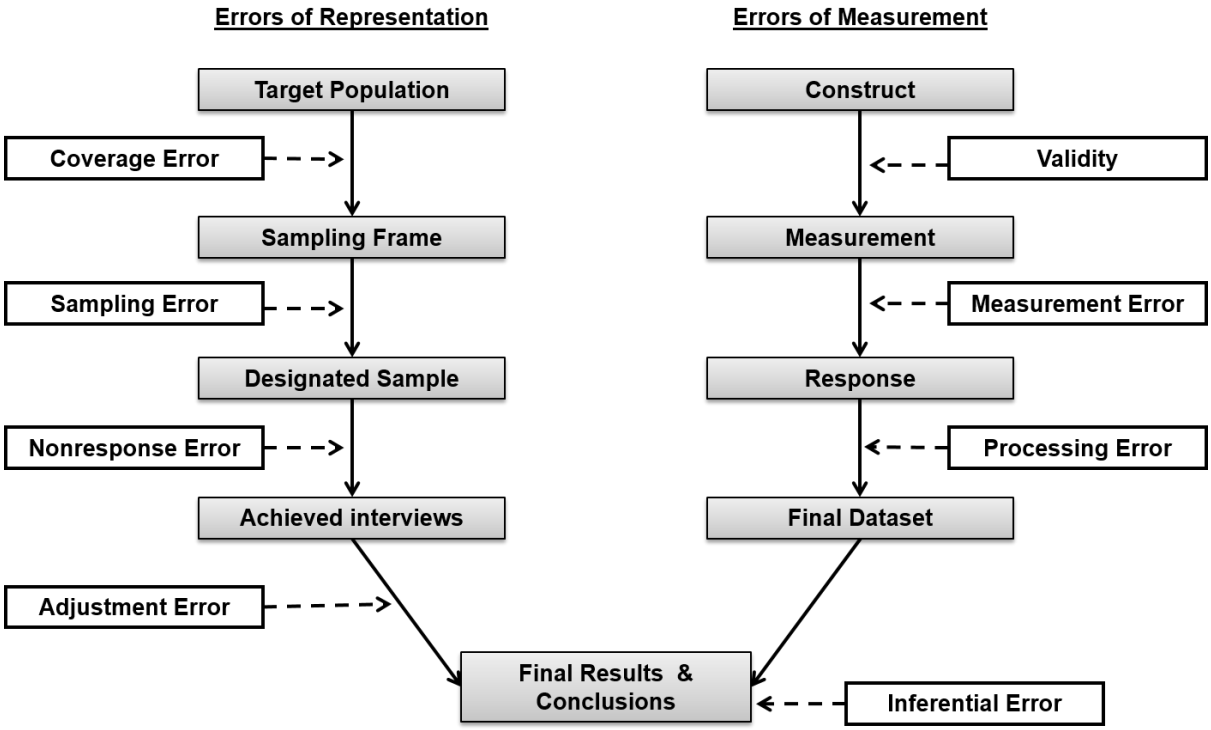
Total Survey Error (TSE) is a component of the Total Survey Quality (TSQ) (Biemer & Lyberg, 2003) concept that focuses on the operational aspects of the survey process. A TSE approach identifies key potential sources of error in the design, collection, processing and analysis of survey data and provides a framework for optimising survey quality within given design and budget parameters. While TSE has been criticised for being an intellectual paradigm than an explanatory, statistical model of survey error (Groves & Lyberg, 2010) it does provide researchers with a strong foundation to assess, reflect on and improve their research practice.

TSE is typically broken down into sampling error, referred to as errors of representation, and non-sampling error, errors of measurement. Errors of representation occur as part of the sample specification and the selection of the cases from the sample frame. Non-sampling error, or errors of measurement, is a much broader concept encompassing systematic and random errors (McNabb, 2014).

Figure 1 on the following page summarises the errors of representation and measurement that can occur at each state of the survey research cycle. The main sources of error affecting survey accuracy include sampling frame errors and omissions (e.g. gaps, biases, inaccuracies in the sampling frame), sampling error (e.g. biases in the respondent selection routine or sub-sampling routines), measurement error (e.g. questionnaire design errors, interviewer errors, respondents errors), non-response error (e.g. both unit-level and item-level non-response) and

data processing errors (e.g. errors in data editing, coding, weighting or the creation of data files or tables) (Biemer, 2010) (Biemer & Lyberg, 2003) (Blausis & Thiessen, 2012) (Groves, M, Crouper, Lepkowski, Singer, & Torangeau, 2009).

Figure 1: Total Survey Error framework in the context of the research cycle.



Adapted by (Lavrakas & Pennay, 2014) from (Groves, M, Crouper, Lepkowski, Singer, & Torangeau, 2009).

Mitigating survey error involves achieving a balance between a survey that meets the desired quality standards and a survey that is cost effective. This need for balance is particularly evident where data collection is on a large scale and requires specialist skills or infrastructure. The resulting tension between quality and expense has the potential to affect all components of the survey cycle. Researchers also need to be mindful that fixing one aspect of survey error can take budget away from other mitigation activities (Blausis & Thiessen, 2012). In addition, it is often the case that addressing one source of error can inadvertently increase

another source of error (Hillygus, 2011). For example, increasing response rates could decrease representativeness.

The intersecting and dependent nature of survey error and its minimisation means that it is essential for institutional researchers to look at errors of representation and errors of measurement from an integrated perspective rather than focusing on and remediating individual components in isolation.

Institutional research and the Total Survey Error framework

To date, neither a TSQ nor a TSE approach has been readily embraced by the institutional research community. This is not to say that survey research in conducted in a unsystematic or inappropriate way within institutions but suggests the absence of an overarching quality assurance and decision making framework. Given that the feedback obtained from institutional surveys provides key sources of data that contribute to organisational intelligence and the 'success' of the educational experience (Borden, Massa, & Milam, 2001) it is critically important that any criticisms regarding the robustness of survey data be addressed so that this information is on an equal quality footing with administrative data.

Liu (2010) presented a detailed conceptual strategy based on TSE for use by institutional researchers undertaking surveys. This framework addressed a useful gap in the literature by linking institutional research practice with a means of holistically investigating and understanding survey data quality issues. Despite this, there is little evidence that the proposed framework has been widely trialled or adopted within an institutional research context.

In the small number of instances where TSE has been used by institutions, the integrated approach to looking at a broad range of survey errors has yielded valuable insights. For example, a common measure of survey quality used by institutional researchers is response

rate. Crow, Johnson & Hanneman (2011) found that when the response rate to a survey of recent graduates was increased by using a multi-mode approach to data collection (phone, email and hard-copy form), representativeness was improved for some demographic variables but weakened for others. While the variables with decreased representativeness weren't critical for the key research questions, this demonstrates the subtle way that something as apparently innocuous as a higher response rate can contribute to another source of survey error.

The University Experience Survey

A consortium commissioned by the Department of Education, Employment and Workplace Relations (DEEWR) designed the UES during 2011. The UES consisted of a survey instrument, the University Experience Questionnaire (UEQ), and a survey methodology (Radloff, Coates, James, & Krause, 2011). It was primarily created to measure levels of engagement and satisfaction of current first and final year undergraduate students at Australian universities. The instrument and survey approach was refined in 2012 by the same consortium.

In 2013 and 2014 Graduate Careers Australia and the Social Research Centre assumed responsibility for contributing to the continuous improvement of the execution of the UES. The UES is currently the largest survey of higher education students in Australia with more than 100,000 students participating each year.

A further dimension to consider in relation to the UES was presented in mid-2014 with the introduction of the Quality Indicators for Learning and Teaching (QILT). The federal budget measure committed to a survey research program aimed at collecting student feedback from undergraduate students, graduates and employers of graduates. As the UES will form the first 'survey plank', supplemented by the Graduate Outcomes Survey (GOS) and the

Employer Satisfaction Survey (ESS), it is essential to ensure that this key component of QILT is as robust and error free as possible prior to the introduction of the new survey elements.

A TSE issue checklist for the UES

The approach to conducting the UES for the 2013 and 2014 cycles was based on a careful consideration of potential sources of survey error tempered by an appreciation of the compressed timeline for both cycles of data collection. As such, it was important to assess and prioritise areas for action and improvement. TSE was used to provide a:

- theoretical and conceptual framework for evaluating the design of the UES,
- structured approach to making decisions about modifying the UES to support continuous improvement,
- method for determining an optimal research design that offered good value for money, and a
- means to challenge accepted paradigms regarding response rate as the primary indicator of a 'good' survey.

The TSE issue checklist on the following page lists examples of the types of questions that were asked with respect to the survey errors that had the potential to affect the UES. The checklist was not intended to be exhaustive but was used primarily to summarise key feedback and observations made in relation to the UES during the 2012 implementation and in the lead up to the 2013 collection. The main issues for mitigation were summarised and an assessment made regarding the extent to which the likely impact on data quality was high, medium or low.

Table 1. TSE issue checklist identifying errors and potential impact on the UES data quality.

	TSE checklist questions	**Issues**	**Potential impact on data quality**
Errors of representation			
Coverage error (under coverage and over coverage)	• How has the in-scope population been defined? • Do the specifications of the sampling frame match the population? • Are there ineligible or duplicated cases in the sampling frame?	In-scope population inaccurately or poorly defined. Sample frame may not be representative of the undergraduate population. Ineligible cases sampled.	High
Sampling error	• Is the sample size appropriate? • Is the margin of error greater than expected?	Sample size inadequate. Data not sufficiently precise for analytic or reporting purposes.	High
Non-response error	• What is the survey level non-response? • Are there any population sub-groups that do not respond to the survey? • What is the item level non-response?	High rates of survey non-response could result in non-response bias. Population sub-groups under represented. High rates of item level non-response could result in non-response bias.	High
Adjustment error	• Is the weighting schema appropriate?	Weighted data may not accurately represent the population.	Low

	TSE checklist questions	Issues	Potential impact on data quality
Errors of measurement			
Validity	• Is the instrument valid? • Is the instrument reliable?	The instrument does not measure the desired concepts or does not measure them consistently.	Low
Measurement error	• Is the questionnaire well designed? • Will interviewers unintentionally or intentionally providing incorrect information?	Poor design leading to inaccurate or incomplete responses or answers that are not relevant to the desired concepts. Interviewers may unintentionally cause respondents to change or modify their responses. Keying errors result from interviewer data input.	Medium
Processing error	• Is the analytic unit appropriately defined? • How will the data be cleaned? • Will the data coded appropriately?	Inaccurate definition of the analytic unit. Inadequate validation checks of outputs. Coding errors or inconsistent coding of open-ended responses.	Medium
Inferential error	• Will the data been analysed and interpret correctly?	Incorrect analytic techniques used. Inaccurate inferences made.	Low

Attempting to address all of the potential survey errors in during one cycle of data collection would be costly and make it difficult to determine which mitigation strategy was effective. For the 2013 UES collection, the main focus was on reducing the errors of representation that were regarded as having the greatest impact on data quality: coverage error and non-response error. In 2014, the error mitigation strategies shifted to address sampling error as well as retaining an emphasis on non-response error. Each of the errors of representation and the errors of measurement that were considered and actioned are discussed in detail in the following sections. Adjustment and inferential errors have not been included as their low risk rating mean that they were not part of the 2013 or 2014 remediation program.

Errors of representation

An initial analysis of the errors of representation relevant to the UES suggested that coverage, sampling and non-response errors had the potential to impact substantially on data quality. Previous recommendations relevant to the UES highlighted particular concerns with the definition of the in-scope population (Radloff et.al 2011, 2012) and the survey response rate.

In the 2013 cycle of the UES, strategies to address coverage errors and survey level non-resposne were prioritised. For 2014, attention was devoted to continuing to mitigate non-response errors while also focusing on sampling error. Coverage error, sampling error, non-response error mitigation approaches are discussed in further detail in the follwing sections. Errors of adjustment, data weighting, have been omitted as they were not considered to be a key issue for the data quality of the 2013 and 2014 UES collections.

Coverage error

For the 2011 trial of the UES and the 2012 collection a ‘bottom-up’ approach to creating the sampling frame was used where institutions provided extracts from their student systems to the contractor based on a narrative definition of the in-scope population. These records were cleaned, coded and formed the foundation of the sample frame.

The 2013 sample frame was based on a 'top-down' approach using population data from HEIMS to create the sample frames for individual universities. This approach minimised accidental bias being introduced in the sample selection process and ensured a nationally consistent approach to sampling. While it would have been ideal to use validated Submission 2 data, due to the timeline for UES data collection.

The success of the 2013 collection, as evidenced by a substantial increase in the participation rate and small proportion of students identifying as out of scope for the survey, indicates that using submission 1 data from HEIMS provides a robust foundation for the sampling strategy. This key change formed the foundation for addressing errors of representation resulting from both over coverage and under coverage of the target population.

Table 2 contains the main coverage errors and the associated mitigation strategies that were employed during the 2013 and the 2014 UES collections. The details relevant to each of these strategies are explored in the following sections.

Table 2. Summary of the identified coverage errors and mitigation strategies used for the 2013 and 2014 UES.

Coverage errors	**Mitigation strategy**
Sample frame not representative of the student population.	Administrative data from the Higher Education Information Management System (HEIMS) data used to generate a national sample frame.
In-scope population poorly defined	Transparent definition of in-scope population created syntactically using established data elements from HEIMS.
Ineligible cases included in the sampling frame	Sample frame independently validated by institutions to ensure that ineligible cases are flagged and unavailable for selection.

Defining the in-scope population. Defining the in-scope population of commencing students was relatively unproblematic given that a data element identifying commencing

students is available in HEIMS. As outlined below, it was challenging to achieve a viable and robust definition of completing students.

Commencing students. For the 2013 and 2014 UES collections, commencing students were defined as first year students who were enrolled in an undergraduate course, studying onshore, commenced study in the relevant target year; and enrolled for at least one semester. This definition was provided to participating institutions in 2012 and relevant records were extracted by the institution and provided to the data collection agency. It is unknown if this definition was operationalised in the same way by each institution.

In 2013 and 2014, records conforming to the agreed definition of a commencing student were extracted from the national HEIMS Submission 1 Student File. Individual institutions were asked to confirm, where possible, that the selected commencing students were still enrolled.

Final year students. For all UES collections, 'completing' students were defined narratively as final year students who were enrolled in an undergraduate course, generally in their third year of study, and studying onshore.

The 2012 definition of final year students noted that these students should have commenced study prior to the target year. This component of the definition was problematic for courses that are 12 months in duration. In 2013, and in the subsequent 2014 collection, students who were enrolled in these shorter courses were included in the sample as completing students.

As was the case for commencing students, in 2012 institutions were responsible for extracting in-scope student records based on this loose definition. The 2012 UES National Report recommended that the definition of completing students be clarified due to confusion at an institutional level regarding which students were in-scope for the collection (Radloff et al, 2012). This confusion appears to have resulted in substantial over coverage of what were thought to be completing students, with 445,332 students identified as in-scope in 2012

compared with 341,343 in 2013. The difference of just over 100,000 students is likely to reflect the inadvertent inclusion of middle years students in the sample frame in 2012.

In 2013, a number of solutions to operationalising the description of 'completing' were trialled as there is no indicator in HEIMS which can be used to identify a final year student. In principle, student progression can be estimated by calculating the ratio of Equivalent Full Time Student Load (EFTSL) completed successfully' and 'currently in progress' to the total EFTSL for the course. Where the ratio was close to 1.0, it was assumed that the student intended to finish the course in the reference year and was therefore a final year student. The main practical difficulties involved in operationalising this definition was collating 'EFTSL completed successfully' by students whose initial enrolment may extend back ten years and estimating what they intended to do in Semester 2 or Summer Term. The task was relatively easy for full-time students in three year courses, but more difficult for part-time and external students, those who took leave of absence or those who transferred from one course to another.

In 2013, two options for identifying final year students were explored with the main difference between the two being that a correction for the duration of the course. This approach using the course length correction seemed to appropriately identify completing the majority of completing students for most institutions. As such, this option was used to identify completing students, with specific adjustments required to accommodate the idiosyncrasies of a small number of universities with less typical course structures.

Specifying the sampling frame. Unless special arrangements are made with individual institutions to survey additional student groups, the UES was restricted to undergraduate students studying at an onshore campus. Although all of the records in the Submission 1 file are included in the sample file to support the verification process, specific groups of students are flagged as excluded from the UES including all students in postgraduate and non-award

courses as well as offshore, international undergraduate students, undergraduate students in the middle years of their course and undergraduates enrolled concurrently.

The sampling frame was clearly specified and defined using HIEMS variables and transparent syntax.

Identification of ineligible cases. All institutions are asked to 'inspect the sample data file for correctness' as the Submission 1 data, unlike the Submission 2 data, are not formally verified. It is also apparent from the inspection of the Submission 1 data that institutions varied enormously in their mid-year intakes, early attrition rates, mid-year completions and the number of commencing students that transfer from one course to another at the end of Semester 1 which had the potential to affect sample quality. As such, survey managers were requested to update student background information that was misreported in Submission 1, not append mid-year intakes to the data file as these students would be excluded from the sample frame and not update the file for the small number of commencing students that managed to affect a mid-year transfer in their first year of study as current course of enrolment was confirmed in the survey.

During the verification stage, institutions are able to exclude students from the survey. The majority of the exclusions relate to course changes that made the student out of scope for the survey (i.e. enrolling in a post-graduate course or deferring). Where possible, information relating to deceased students was removed from the sample files. There was no indication as part of the 2013 or the 2014 UES that any institution made inappropriate exclusions and in most cases, all of the selections identified during the sampling process were retained by the university.

Sampling error

Issues relevant to sampling error were not specifically addressed in 2013 as there was insufficient information about the operational aspects of the 2012 collection to implement a

remediation strategy. It was also potentially pointless to fix perceived sampling issues while the concerns regarding the sampling frame were being addressed.

The mitigation strategies listed in Table 3 were linked and dependent meaning that it was not possible to disentangle errors of sample size appropriateness from expected margin of error. As such the following section provides an integrated discussion of sample sizes, margins of error and anticipated response rates.

Table 3. Summary of the identified sampling errors and mitigation strategies used for the 2014 UES.

Sampling errors	**Mitigation strategy**
Sample size inappropriate	Implement an approach that supports accurate sampling at a strata level (rather than nationally)
Margin of error higher than expected	Moderate expectations regarding precision of estimates

Sample size inappropriate / margin of error higher than expected. The approach used to determine the appropriate sample size was broadly consistent across the 2012 and the 2013 UES collections. Commencing and 'final year' students were separately allocated to one of 45 Subject Areas used for reporting purposes on the department's website. The subject area and status (commencing or final year) groups were referred to as 'strata'. Using the assumptions outlined in the 2012 UES National Report, all eligible students were selected for strata with up to 1,333 students, effectively a census of those strata. With larger strata, a random sample of 1,333 students was drawn in the hope that this would yield at least 200 responses. The value of 200 was derived from a desire for error bands of ±5 per cent at a 95 per cent level of confidence.

An analysis of this approach in 2013 suggested that this approach to determining the sample size had a number of shortcomings. In general, large strata were substantially over sampled

and often achieved completed interviews well in excess of the target of 200. As the targets for each strata were uncapped, with all students invited to take part were able to complete the survey, students from these large strata were substantially over represented. This has the flow on effect of increasing the gender imbalance as many of the large strata consisted of course offering where males are traditionally under represented such as nursing and education. Lastly, the sampling approach did not take into account the wide range of differential response rates across strata.

In 2014, required sample sizes were calculated at the strata level taking into account the number of records available for sampling and the requirement to report data at a 90% confidence level, +/- 5%. A finite population correction was also applied to each stratum. Using this approach to sample size identification, it was apparent that when the required sample size was compared with the response rates achieved in 2013, it would not be possible to achieve the required number of interviews for a substantial proportion of the strata. This was primarily due to the fact that for many institutions, only a small number of courses are offered in each subject area. When the 2014 sampling approach was applied, the majority of the strata were a census, rather than a sample and a response rate of up to 100 per cent was required in order to meet the required level of reporting precision for an untenably large proportion of the strata.

In consultation with the department, the level of reporting precision was modified to a 90% confidence level +/- 7.5%. In almost all instances, the number of records that needed to be sampled was retained but the required response rate was lowered to a level that was more achievable. It was still the intention of the operational team to aim for a 5% confidence interval and this was used as a 'background target' with a view to making this the actual target in future implementations of the UES.

In both 2013 and 2014, the sample selection was checked against population parameters to confirm that appropriate proportions of gender, qualification, mode of attendance, broad field of education and citizenship characteristics were present in the sample.

Non-response error

Non-response was a clear area for improvement with high levels of survey level non-response, sub-group under representation and item-level non-response reported for the 2012 UES. Table 4 identifies the specific strategies that were implemented in 2013 and 2014 to address these survey errors.

Table 4. Summary of the identified non-response errors and mitigation strategies used for the 2013 and 2014 UES.

Non-response error	**Mitigation strategy**
Survey non-response unacceptably high	Collaborative relationships established with survey managers. Appropriate incentivisation scheme implemented. Response rates monitored and corrective action taken throughout fieldwork.
Population sub-groups underrepresented	Sample representativeness monitored and corrective action taken through targeted reminder emails and SMS' throughout the fieldwork period.
Item level non-response unacceptably high	The input controls for each item part of the questionnaire logic programmed into the survey.

The following sections detail the approaches undertaken to mitigate survey non-response, population sub-group representation and item non-response for the 2013 and 2014 UES.

Survey level response. For the 2012 UES, the main focus for response maximisation activities was on achieving the 35 per cent overall response rate. Thirty-five per cent was regarded as a 'feasible stretch target' appropriate for census level surveys (Radloff, et al 2011). The online survey conducted in 2012 fell well short of this target, only achieving a 20.2 per cent response rate and necessitate the deployment of a telephone non-response follow-up survey which introduced mode and interviewer effects (Radloff, et al 2012).

Features of both the data collecting mode and the interviewers have the potential to increase measurement error and contribute to TSE.

In 2013, a consistent 'whole-of-UES' response maximisation strategy was deployed which aim to make substantial improvements to the online response rate so that telephone interviewing would not be required. This strategy included the use of incentives for each institution, a follow-up, hard copy letter for students who did not respond to the original email invitation and generic and targeted email reminders. Incentives were allocated on an institutional basis, with $1,000 worth of prizes being drawn for each institution. The incentives consisted of a major prize per institution of a Coles Myer gift voucher worth $500 and five runners up prizes of $100 Coles Myer gift vouchers. The prize draws were timed to encourage prompt online completion, typically within the first two to three weeks of survey deployment.

Students were sent one initial invitation email and up to five email reminders. The emails were UES-branded, html-enabled and included a hyperlink directly to the online survey as well as manual login and helpdesk details. Students were able to advise of a change to their enrolment status, 'opt out' or unsubscribe by reply by of email. The standard email reminder schedule was four, 11, 16 business days and 20 business days following the initial email invitation.

The second email reminder was timed to arrive shortly before the prize draw cut-off. Email reminders three and four used tailored text to target specific groups of students with low response rates. The majority of institutions also employed a fifth, targeted email reminder before survey close.

Table 5 on the following page illustrates the substantial improvement in the national response rate from 2012 to 2013 with an overall increase of just over nine percentage points. With the exception of two universities that had very good response rates in 2012, all institutions

demonstrated substantial response rate improvements of up to 26 percentage points. In contrast, the response rate decline was, at most, 0.3 percentage points.

Table 5. Institutional, online response rates to the 2012 and 2013 UES.

	2013	**2012**	**Change**
Australian Catholic University	23.7	11.6	12.1
Bond University	32.8	6.7	26.1
Central Queensland University	36.0	25.3	10.7
Charles Darwin University	40.5	25.7	14.8
Charles Sturt University	32.3	21.7	10.6
Curtin University of Technology	26.1	23.8	2.3
Deakin University	29.2	14.7	14.5
Edith Cowan University	29.3	25.7	3.6
Flinders University	35.2	21.1	14.1
Griffith University	23.5	19.5	4.0
James Cook University	29.0	19.1	9.9
La Trobe University	33.0	20.7	12.3
Macquarie University	26.3	18.9	7.4
MCD University of Divinity	50.5	44.6	5.9
Monash University	39.7	23.3	16.4
Murdoch University	30.6	20.1	10.5
Queensland University of Technology	29.4	20.8	8.6
RMIT University	20.8	3.2	17.6
Southern Cross University	24.4	15.3	9.1
Swinburne University of Technology	25.5	13.2	12.3
The Australian National University	29.3	29.6	-0.3
The University of Adelaide	41.4	24.6	16.8
The University of Melbourne	34.5	22.0	12.5
The University of Notre Dame	26.0	17.1	8.9
The University of Queensland	32.5	24.9	7.6
The University of Sydney	30.3	23.1	7.2
The University of Western Australia	39.7	39.8	-0.1
University of Ballarat	22.1	20.4	1.7
University of Canberra	24.4	19.8	4.6
University of New England	32.9	16.3	16.6
University of New South Wales	27.0	17.5	9.5
University of Newcastle	34.0	30.9	3.1
University of South Australia	25.2	23.6	1.6
University of Southern Queensland	25.2	15.7	9.5
University of Tasmania	33.0	22.7	10.3
University of Technology Sydney	28.2	13.4	14.8
University of the Sunshine Coast	29.2	23.5	5.7

University of Western Sydney	26.6	22.2	4.4
University of Wollongong	23.5	20.1	3.4
Victoria University	17.9	10.4	7.5
Total	**29.3**	**20.2**	**9.1**

Feedback from survey managers and observations of the online rate of return across a number of institutions suggests that competing survey activity, including recent data collection activities with students, negatively impacts on the response rates. Uncertainty around when the UES would be undertaken each year resulted in reluctance by some universities to commit to reducing the number of student surveys undertaken in the latter half of the year.

The certainty offered to institutions by the QILT initiative will allow them to plan for the UES and to embed this instrument within their survey calendar. This confidence in the department's commitment to the UES will also assist with the creation of 'clear space' in each university survey schedule which is so important to achieving robust response rates. This forward planning of future survey activity will be a key feature of the 2015 UES.

Population sub-group representation. Table 6 on the following page shows that there are a number of sample parameters that closely match the achieved respondent profile. Status, course of study, course of study type, ATSI status, and type of attendance are similar for both sample members and survey respondents. Language spoken at home and citizenship status indicators are also surprisingly similar for the sample and the achieved respondent profile given that students with these characteristics are traditionally less likely to participate in similar surveys, such as the Australian Graduate Survey.

As was the case with respect to the 2012 UES (Radloff, et al 2012), the largest potential source of survey level non-response bias was in relation to gender with male students exhibiting substantially lower response rates overall than female students. For the 2012 survey, the proportion of male online respondents was 35.1%, similar to the 33.3% achieved in 2013. While this level of male non-response is still not optimal, at least the large increase

in overall response rate for 2013 did not result in a decrease in gender representativeness of the same magnitude.

Representativeness is one of the key areas targeted for improvement in the 2014 UES.

Table 6. Comparison of the sample and respondent characteristics for the 2013 UES.

	Sample	%	Respondents	%
Base	344,692		100,225	
Status				
Commencing	208,307	60.4	59,653	59.5
Final year	136,385	39.6	40,572	40.5
Gender				
Male	148,264	43.0	33,349	33.3
Female	196,428	57.0	66,876	66.7
Combined course of study indicator				
Combined/double degree	37,887	11.0	11,919	11.9
Single degree	306,805	89.0	88,306	88.1
Course of study type				
Bachelors Graduate Entry	4,925	1.4	1,627	1.6
Bachelors Honours	10,096	2.9	3,921	3.9
Bachelors Pass	320,155	92.9	92,808	92.6
Associate degree	4,959	1.4	908	0.9
Advanced Diploma	1,494	0.4	408	0.4
Diploma	2,811	0.8	495	0.5
Other undergraduate award	252	0.1	58	0.1
Aboriginal and Torres Strait Islander				
Indigenous	4,126	1.2	1,067	1.1
Non-Indigenous	334,617	97.1	97,535	97.3
Not stated	5,949	1.7	1,623	1.6
Type of attendance code				
Full-time	307,739	89.3	90,137	89.9
Part-time	36,953	10.7	10,088	10.1
Language spoken at home code				
English	258,416	75.0	77,208	77.0
Language other than English	81,537	23.7	21,931	21.9
Not stated	4,739	1.4	1,086	1.1
Citizen/resident indicator				
Domestic	294,662	85.5	88,067	87.9
International	50,030	14.5	12,158	12.1

Item level non-response. Item level non-response for the 2013 UES was compared with the online component of the 2012 UEQ. Average item non-response to the 2013 survey was 1.0 per cent, a substantial reduction from 7.7 per cent average non-response to the 2012 online survey. Those students who responded to the UES in relation to a second course exhibited slightly higher levels of item level non-response, 1.6 per cent on average, which is understandable given that the survey was substantially longer for this group.

An examination of some of the items with comparatively higher levels of non-response suggests that the wording of the questions is awkward which may be contributing to the students declining to answer. Taking acdavail and acdhelp (below) as examples (1.4 per cent and 1.5 per cent non-response respectively) the question stems and response frames are not as compatible as they could be.

acdintro During 2013, to what extent have you found <u>academic or learning advisors</u> to be…

STATEMENTS

acdavail (36) Available?
acdhelp (37) Helpful?

RESPONSE FRAME

1. Had no contact
2. Not at all
3. Very little
4. Some
5. Quite a bit
6. Very much

It is difficult to determine what 'quite a bit available' or 'some helpful' actually means. Cognitive testing of these items or a revised response frame could further reduce the non-response to these UEQ items. The UEQ is scheduled for review prior to the next cycle to ensure that it aligns harmoniously with the Graduate Outcome Questionnaire and the Employers Satisfaction Questionnaire that are being developed and refined in 2014/15.

Errors of measurement

Errors of measurement associated with the UES were not the main priority for correction during 2013 or 2014 collections as, at worst, they were regarded as presenting medium levels of threat to data quality. As inferential error was seen to be a low risk, no steps were explicitly taken to counter inferential concerns. Based on a review of the instrument prior to fieldwork in 2013, validity was monitored but no remedial action was taken in 2013 or 2014. The following sections identify errors of measurement that were considered, particularly in relation to measurement and processing error and the mitigation strategies that were implemented.

Validity

Substantial effort was involved in the initial development and testing of the UEQ in 2011. As such, these activities were not repeated during the first full deployment of the survey in 2012, largely due to the fact that there did not appear to be any issues of concern relating to the instrument.

During the initial survey set-up procedures associated with the 2013 UES, the department was provided with extensive feedback regarding the wording, sequencing and structure of the UEQ. This review of the instrument was undertaken from an operational perspective with a view to minimising measurement error. Due to the time constraints associated with the deployment of the survey in both 2013 and 2014, any modifications to the survey, aside from those required to collect data at the course or program level, were held over to be considered for future implementations of the UEQ. It was also the case in 2013 that a number of higher priority improvements were made to minimise survey error associated with the UES and it was more prudent to evaluate the efficacy of these modification before addressing lower priority issues.

Even a cursory inspection of the instrument identifies issues with the face validity of the items, specifically with respect to the use of technology and online study modes. As noted below in relation to measurement error, the wording of some items and response frames could be contributing to elevated item level non-response for a select number of questions. Consideration will be given to reviewing the instrument prior to the 2015 UEQ, particularly in light of the creation of the GOS and the refinement of the ESS.

Measurement error

Measurement error was identified as a lower priority for mitigation when compared with compared with the higher priority concerns related to errors of representation. Table 7 shows that a substantial change to the UES regarding the mode of data collection was implemented as a mitigation strategy. Altering the data collection approach from mixed-mode to single mode represented a substantial modification to the research design and, as such, additional mitigation activities were not undertaken in 2013 or 2014 to address, for example, to refine the questionnaire.

Table 7. Summary of the identified measurement errors and mitigation strategies used for the 2013 and 2014 UES.

Measurement error	**Mitigation strategy**
Questionnaire design inadequate	-
Interviewers elicit biased responses or enter data incorrectly	Undertake the UES in a self administered mode.

Questionnaire design. As the questionnaire had been extensively tested during the development phase and the initial 2012 collection, only minor modifications were made in 2013 as mentioned in the previous section on validity. In 2014, a change was made to the order in which the UEQ modules were presented to students. The 2013 UES displayed the

modules in a number of different orders, referred to as rotations. Order effects were not found to not have a substantial effect on key UES metrics and for the 2014 collection, the module order was fixed in an intuitive logic flow to improve the respondent experience. These alterations were regarded as questionnaire maintenance and not mitigation activities.

As discussed in the section on item-level non-response, there appears to be some indication that questionnaire design could be contributing to incomplete responses. Additional testing of the UEQ will be undertaken, if practicable, prior to the 2015 deployment of the UES to assess whether questionnaire design issues could be contributing to survey error.

Interviewer related error. In 2012, a mixed mode approach to data collection was used primarily as a non-response maximisation strategy when the online survey failed to yield a useful number of completed surveys. Telephone interviewing was undertaken at a select number of universities suggesting that the main purpose of this exercise was to maximise returns and not to improve representativeness.

The single mode of data collection was introduced in 2013 primarily to improve fieldwork efficiency. It was not a key aim to improve measurement error by restricting data collection to one mode but it did helpfully eliminate any potential mode or interviewer effects that may have been present.

Processing error

Potential processing error was initially identified as medium risk for the 2012 UES, largely because the potential scope of the error was unknown and it was more conservative to monitor the data processing approach than regard the error potential as low. Table 8 identifies the risk mitigation strategies that were used to effectively minimise error for the 2012 and 2013 UES.

Table 8. Summary of the identified processing errors and mitigation strategies used for the 2013 and 2014 UES.

Processing error	**Mitigation strategy**
Analytic unit inaccurately defined	Clear definition of the analytic unit created syntactically.
Open ended responses coded inappropriately	Senior, experienced coders responsible for coding activities. Detailed, structured procedures for data cleaning. Independent checking of ten per cent of the coding workload. Verification of queries by individual institutions.
Data cleaned inaccurately	Input controls minimising the amount of cleaning required. Detailed, structured procedures for data cleaning. Independent verification of cleaned data by two members of the analytic team.

Definition of the analytic unit. The analytic unit for the 2012 UES was the student and the data file contained one record for each respondent to the survey. For the 2013 UES, the changes to the instrument to allow students in double degrees to respond separately in relation to each degree required that the analytic unit be altered to the course. In the 2013 data set, a UEQ was defined as valid and complete if

- the student had completed units in the course
- there was a minimum of one valid UEQ scale score, and
- in the case of double degrees for which the student had at least one valid UEQ scale score for each course, the courses were in different subject areas.

Where double degree students had completed units in both degree components and they were in the same subject area, the first record was selected for analysis. The definition of the analytic unit was operationalised syntactically to create uniformity across institutions.

Coding of open ended responses. To ensure consistency in the cleaning process, records were first merged from all separate institution level datasets (as collected on the online platform) into one master dataset.

A master course list for each institution was used to look up new course names entered by students. In those cases where a course name matched multiple course codes the record was assigned to the course with the highest enrolment. Where an appropriate course code for the course name or double degree recorded by the student could not be found, queries were sent to the institutional survey manager for resolution.

Of the responses coded, several broad categories of anomalous response requiring further editing were identified including the student entering an invalid course title, a course not offered by the institution or reporting that they were enrolled in a double degree but recording the same course title for both degrees. This respondent information will be used to asses whether changes need to be made to the questionnaire to reduce input errors.

Data cleaning. As was the case with respect to data coding, to ensure consistency in the cleaning process, records were first merged from all separate institution level datasets into a master dataset. Records with newly entered course information were edited, and final course level, field of education and subject area information was derived from a master course list based on available course data for each institution. Where new course codes were added to the master course list, accompanying information was sourced from the survey manager for the relevant institution.

The in-scope status of the student, that is whether they are enrolled in a degree eligible for the UES, was then re-derived based on revised course level data, to exclude students that had switched from an eligible undergraduate course to an ineligible course such as postgraduate coursework or research. All items in the body of the questionnaire were re-filtered to their

respective bases to ensure there were no errant responses. Items applicable to the student for which data was missing were assigned code 99 ("Item skipped" in the data dictionary).

The entire cleaning process was undertaken in R and independently replicated in SPSS to allow for the early identification and resolution of any error introduced during data editing and cleaning.

Overall quality improvement & cost effectiveness

During 2013 and 2014, many of the pre-existing survey errors associated with the UES were specifically addressed and minimised. Table 9 summarises the errors addressed, the key drivers of improved quality and the assessed change in the risk ratings. With the exception of validity, which wasn't explicitly included in the survey error improvement program, all types of errors identified for mitigation were reduced. The greatest improvement was made in relation to coverage error and while there is still some room for further improvement, for example by tweaking the algorithm used to identify completing students, there is little value to be gained by continue to devote intense effort to over coverage or under coverage of the in-scope population.

Sampling error, non-response error and validity still potentially present some risk to the data quality of the UES and these areas will be targeted for improvement in 2015. Given the nature of the quota groups and the requirement for substantial response rates for those institutions that have smaller enrolments but a large number of subject areas, this will continue to be a challenge for future iterations of the UES. Validity was not explicitly addressed in 2013 or 2014 but there is some evidence to suggest that it would be appropriate to review this potential contributor to survey error. As such, the risk rating has increased for validity from low to medium.

Table 9. Key drivers of improved quality of the UES and risk ratings for survey error.

Type of error	Key drivers of improved quality	Original risk rating	Current risk rating
Coverage error	Use of HEIMS, robust definition of 'completing' students, rigorous institutional validation process	High	Low
Sampling error	Improved response rates, refined sampling strategy, fine-level quota control of study areas within institutions	High	Medium
Non-response error	Increased unit level response rates, reduced item level non-response, and stable under representation of some sub-groups.	High	Medium
Validity	-	Low	Medium
Measurement error	Use of a single mode of data collection and data input logic controls.	Medium	Low
Processing error	Clear definition of analytic unit, documented data cleaning and file preparation procedure, well-developed quality checks for coding and an effective query resolution process	Medium	Low

One of the useful features of TSE is the appreciation of the desire to achieve the best survey quality outcomes within a, typically constrained, budget. Essentially, the cost of implementing the error mitigation strategies for the 2013 and 2014 UES had a neutral impact on the overall budget when compared with the 2012 cycle of data collection. The money saved by just using a single mode of data collection instead of supplementing online responses with telephone interviews was used to undertake the response maximisation activities required to reduce the unacceptably high survey level non-response. Decreased survey error in relation to coverage error, sampling error and processing error resulted from the use of established ISO procedures and some good ideas about the way pre-existing data and resources, such as HEIMS, could be used creatively and effectively.

It would also be remiss not to acknowledge the contribution of the survey managers and fellow institutional researchers who actively supported, tested and engaged in the survey and error mitigation process. The cost of their time has not been taken into account but has undoubtedly contributed to the reduction of a range of error sources, particularly during 2013 when coverage and non-response errors were being addressed with vigour.

Implications for Institutional Research

The implementation of the TSE framework described using the UES as a case study provides a worked example of the types of issues that arise and need to be taken into account when designing and implementing a large-scale survey program. It is possible that some institutional survey managers can become complacent when undertaking survey research as they are typically executing a census of a population from a sample frame that they have direct access to via a student administration system. While this may reduce the need to focus on sampling related errors, there are still a number of survey errors that could be present. The analysis of the threats to the quality of the UES data clearly shows both errors of representation and errors of measurement are present in a survey program that has been rigorously tested and developed, highlighting the need for the monitoring and continuous improvement of survey designs and processes.

In general, institutional researchers planning to undertake or review survey programs could benefit from:

- using TSE as a conceptual framework to provide a basis for identifying and classifying relevant survey errors, and
- undertaking a risk assessment of the identified surveys errors provides a focus for determining the areas of concern to be prioritised.

A formal assessment of the survey error and the associated risks to subsequent data quality also provides the basis to evaluate whether or not planned mitigation strategies have the potential to be good value for money. For example, it may be worthwhile to implement an expensive response maximisation strategy if high levels of survey non-response are likely to have a substantial impact on data quality.

Lastly, having a clear understanding of the survey errors and the associated threat to data quality supports an additional approach to prioritising mitigation activities. Errors that are earlier in the research cycle can flow through to later data collection activities so it may be worthwhile considering rectifying these errors, even if they are a lower priority, before moving onto 'downstream' issues. It may be the case, for example, that undertaking additional surveys in another data collection mode to improve representativeness may not reduce the overall survey error if sampling errors have not been addressed.

TSE is obviously not a 'magic bullet' for improving institutional research survey quality but it does provide a clear foundation for evidence based decision making about the quality of survey design, implementation and the resulting data outputs. It is also a useful means for institutional researchers to continuously improve their practice with respect to survey research.

Bibliography

Biemer, P. (2010). Total Survey Error: Design, Implementation and Evagluation. *Public Opinion Quarterly, 74*(5), 817-848.

Biemer, P., & Lyberg, L. (2003). *Introduction to Survey Quality.* USA: John Wiley & Sons.

Blausis, J., & Thiessen, V. (2012). *Assessing the Quality of Survey Data.* USA: Sage Publications.

Borden, V., Massa, T., & Milam, J. (2001). Technology and tools for institutional research. In R. (. Howard, *Institutional Research: Decision Support in Higher Education* (pp. 201-209). Tallahassee, Florida.

Crow, D., Johnson, M., & Hanneman, R. (2011). Benefits and Costs of Multi-Mode Survey of Recent College Graduates. *Survey Practice, 4*(5), 1-6.

Groves, R., & Lyberg, L. (2010). Total Survey Error: Past, present and future. *Public Opinion Quarterly, 74*(5), 849-879.

Groves, R., M, F. F., Crouper, M., Lepkowski, J., Singer, E., & Torangeau, R. (2009). *Survey Methodology.* New Jersey: Wiley.

Hillygus, S. (2011). The Practice of Survey Research: Changes and Challenges. In A. Berinsky, *New Directions in Public Opinion* (pp. 32-51). Routledge.

Lavrakas, P., & Pennay, D. (2014, 3 21). *QPMR Webinar: Total Survey Error.* Retrieved 9 19, 2014, from The Social Research Centre: http://www.srcentre.com.au/news-events/news-and-events-item/2014/03/21/qpmr-webinar-total-survey-error

Liu, Q. (2010). A Survey Data Quality Strategy: The Institutional Research Perspective. *Association for Institutional Research,* (pp. 1-34). Atlanta, Georgia.

McNabb, D. (2014). *Nonsampling error in social surveys.* USA: Sage.

Radloff, A., Coates, H., James, R., & Krause, K. (2011). *Report on the Development of the University Experience Survey.* Canberra: Australian Government.

Radloff, A., Coates, H., Taylor, R., James, R., & Krause, K. (2012). *2012 University Experiene Survey National Report.* Canberra: Australian Government.