

Methods in feedback research

Gavin T L Brown

The University of Auckland

Lois R Harris

Central Queensland University

Feedback is meant to inform and guide improvement (Hattie & Timperley, 2007). Researchers have been striving to understand how feedback can be optimised so it is of maximum value to learners, leading them to adaptive and growth-orientated action (Lipnevich, Berg, & Smith, 2016). Understandably, multiple methods of research have been utilised and this diversity of approaches is encouraged since each method has its own strengths and limitations. However, as Guénette (2007) notes, part of the problem in answering questions about feedback efficacy is that studies suffer from design inconsistencies. This has significant impact on review studies and meta-analyses which seek to make sense of the field because the variability in results in those reviews may arise from the variability of methods used in contributing studies. Because feedback studies use differing methods of data collection and analysis, these design differences lead to divergent results and conclusions. Thus, some of what we claim to know about feedback may be a consequence of researchers' methodological decisions, making it essential that not only are studies carefully designed but that comparability is increased.

This chapter surveys the different methods currently being used to investigate feedback in empirical studies that examine real-world processes in the design, delivery, and impact of feedback. While grounded in our own almost decade of work on the topic, this chapter is not an encyclopaedic review of all studies having to do with feedback. Rather we have surveyed the literature to identify common methods being used and for each have reported a few studies that make use of it to illustrate how it is being applied. Readers will no doubt know of others that we have not included; the goal within this chapter is simply to identify methods being used and describe some strengths and limitations for each, allowing readers to make informed methodological decisions in their own research endeavours and adopt a more critical stance toward the research that they read.

As the focus is on empirical work, we have excluded many excellent sources that are conceptual or theoretical analyses of feedback, along with the many literature reviews so as to concentrate on the methodological choices made in individual studies. Additionally, given the chapter's focus is on feedback designed to directly enhance student learning, we have also excluded studies of student feedback on or evaluations of courses, teaching, or teachers since such studies do not examine how feedback is related to student learning; instead such studies are more properly considered institutional evaluation (e.g., Kember, Leung, & Kwan, 2002; Richardson, 2005). This exclusion includes reviews of departments or programs within institutions used as feedback for self-improvement (e.g. Maier, 2010; Schildkamp, & Visscher, 2010).

We have structured this chapter first around methods used for collecting data within feedback studies and second around methods used for analysis of these data. Especially under data collection, we have taken an approach that links quantitative and qualitative methods such that the mixing of methods to overcome method effects is prioritised. This means that we are not methodological purists and in this chapter are trying to prioritise a pragmatic approach.

Data collection

We have grouped data collection methods in a descending order of researcher control. Experimental studies achieve control through random assignment of participants into treatment or control conditions and all materials and activities administered in the treatment group are controlled by the researcher. Slightly less control occurs in quasi-experimental studies where naturally occurring groups or units are assigned to experimental treatments or not. Researchers still have control over data collection when they stimulate participants with highly structured processes such as closed-response questionnaires or stimulated recall of previous events. Free response mechanisms (e.g., focus groups, semi-structured interviews, or draw-a-picture) techniques give the researcher control over the general content of the data, while participants are able to freely respond as they see fit. Observation of naturally occurring data (e.g., video-records of classroom activities or traces in course-books) may reflect the pinnacle of ecological validity because researchers do not prompt or solicit such data, but much data may not be relevant to the empirical goals of the study.

Experiments

Experimental studies exercise the greatest control over conditions and participants by randomly assigning members of a sample, derived from a population, to one or more conditions that experience different processes (Campbell & Stanley, 1963). It is presumed in the random assignment process that differences between individuals are cancelled out in the group membership; that is, the average ability, attitude, or background variables are equivalent between treatment and control conditions. Feedback experiments generally either control the content of feedback or the processes of how feedback is implemented. A few studies have actually developed interventions and experimentally tested whether learners can be taught to provide better feedback.

As an example of the kind of control exhibited in experimental studies, consider the design of Ashwell's (2000) experiment where he manipulated the content of feedback foreign language students received on drafts one and two of a composition. Students were randomly assigned into four conditions. The conditions were: (1) content feedback (draft 1) followed by form feedback (draft 2); (2) form (draft 1) before content (draft 2); (3) mixed form and content feedback (drafts 1 and 2); and (4) no feedback. Statistical analysis of data indicated that there were no significant differences between the four conditions in relation to changes in content. Students receiving feedback did improve their accuracy significantly more than those in the control group; however, the particular pattern of feedback they received did not appear to matter.

Other examples of studies that have manipulated different kinds of feedback include: Butler, Godbole, and Marsh, 2013; Chase and Houmanfar, 2009; Harks, Rakoczy, Hattie, Besser, and Klieme, 2014; Narciss et al., 2014; and Pekrun, Cusack, Murayama, Elliot, and Thomas, 2014. Experimental studies that have examined the process of feedback delivery

include: Bayerlein, 2014; Cho and MacArthur, 2010; Corbalan, Kester, and van Merriënboer, 2009; Hamer, Purchase, Luxton-Reilly, and Denny, 2015; Harber, Gorman, Gengaro, Butisingh, Tsang, and Ouellete, 2012; Lipnevich and Smith, 2009b. Some studies that have mixed process and content factors of feedback including: Strijbos, Narciss, and Dünnebier, 2010 and Vollmeyer and Rheinberg, 2005. Perhaps on the edge of future research is the experimental study that examined, using *fMRI*, the brain activity of participants when given bogus performance information within the neuro-scanner (Kim, Lee, Chung, & Bong, 2010). This study found that brain regions associated with negative affect were activated when norm-referenced feedback was given to low-competence participants, while the same region was activated when criterion-referenced feedback was given to high-competence participants. Further, they found that performance-approach goal scores correlated positively with activation in the negative emotion brain areas during norm-referenced feedback. Although negative emotions can be activating (Pekrun, Goetz, Titz, & Perry, 2002), the authors use these results to caution against the use of norm-referenced feedback. Clearly, further research into this matter is required.

These studies allow reasonably strong claims to be made as to which factors have the greatest impact through the power of the random assignment of reasonably large to large samples of participants into a variety of controlled conditions. Replication studies and use of multiple dependent measures also contribute to the strength of such claims. Such highly-controlled designs are often accompanied by sophisticated analytic processes to establish the impact of demographic variables as moderating or mediating variables (path or structural equation modeling), and evaluation of mean differences (multiple analysis of variance or multiple regression).

A critique of such research is that results in a controlled study do not necessarily generalise to natural environments (e.g., classrooms) where multiple factors simultaneously impinge upon the planned treatment, which itself may not necessarily be implemented with fidelity. This contrast between ‘in vitro’ experimental studies and ‘in vivo’ natural environments can mean that results do not translate into the real world (Autoimmunity Research Foundation, 2012; Zumbo, 2015). Even when experimental studies do take place in natural classroom environments, it is virtually impossible to control for all differences as there are so many potential points of variation, even between hypothetically comparable class groups (e.g., enthusiasm of the teacher; quality of classroom relationships between teachers, students). A further complication arises in the ethics of some students not receiving a treatment that is believed to benefit all; variations such as delayed or counter-balanced treatments need to be considered.

A second important issue with these studies is that, from an inspection of the listed studies, few are true replication studies in which the same methods and variables are used. This means that an important aspect of modern science is missing—replication of experiments demonstrates the robustness of results (National Research Council, 2002; Simons, 2014). Although the frequency of replications is rising (Makel, Plucker, & Hegarty, 2012), there is still little academic incentive to conduct or publish replication studies (Earp, & Trafimow, 2015; Neuliep, 1990), and arguably too few occur to achieve accurate measurement of effect (Hunter, 2001). These concerns are raised in the discipline of psychology, while levels of concern about the same issue in the field of education are

somewhat less prominent. Hence, there is a need to substantiate that findings reported as statistically significant are in fact significant and not artefacts of chance (i.e., all statistically significant results could be the 'chance' event) and that they may legitimately be generalisable to other similar environments. Successful replications seem to occur most often when the original researcher is kept in the research team (Makel, Plucker, & Hegarty, 2012). Obviously, for any kind of replication to be possible, a highly detailed explanation of procedures must be included in reports. Standards as to how experimental studies need to be conducted should be considered by all seeking to use this method (Lindsay, 2015). Nonetheless, the existence of meta-analytic summaries around psychological processes, including feedback (Kluger & DeNisi, 2006; Hattie & Timperley, 2007), does support the idea that sufficient information about how a phenomenon has been studied exists to allow replicable study, even if the findings differ; otherwise, there would not be sufficient comparable studies of a phenomenon to reach any conclusions about the robustness of effect (Smith, Smith, & Smith, 2017).

Quasi-experiments

Similar to the experiment is the quasi-experiment in which naturally occurring samples are randomly assigned to controlled conditions (Shadish & Luellen, 2006). These studies exercise control over the intervention but not over who is within each group, meaning that differences between groups may pre-exist that can impact upon results. Normally, techniques like analysis with covariates can help to account for such pre-existing differences.

A strength of quasi-experimental feedback studies is that they are more likely to take place within classrooms, meaning they may more accurately emulate real world feedback conditions than laboratory experiments. For example, in Gielen, Peeters, Dochy, Onghena, and Struyven's (2010) quasi-experimental study, they adopted a repeated measure design to examine if peer feedback was more effective if the student had to reflect on and write a response to the peer feedback or not. The study drew on 43 Year 7 (6th grade) students all taught by the same teacher, but belonging to two different class groupings, and found no difference in the quality of the final student products based on the two conditions. Another example is Gan and Hattie's (2014) quasi-experimental intervention study where one naturally occurring group of students was taught how to give feedback with a parallel group acting as control. The study found that prompted peer feedback had a significant effect on the number of comments students made related to Knowledge of errors, Suggestions for improvement, and Process level feedback. Clearly as such studies have great potential for better identifying what may effective within in vivo classroom situations, it is hoped that more studies adopt such a design in the future as it is currently one of the least represented methods.

Non-experimental Methods

A wide variety of non-experimental data collection techniques has been used in the study of feedback. These include interviews (including stimulated recall), focus groups, surveys, free response, observations, collection of traces, and case studies. Under this label, we include studies which have a treatment, but no control group. For example, similar to an experimental study of not reporting grades (Lipnevich & Smith, 2009b), Sendziuk, (2010) reported, without a control group, feedback without grades being given to students who then

self-assessed their own writing, with subsequent statistical comparison of peer and tutor marking.

Although non-experimental studies lack random assignment and control groups, they can vary in the degree of control researchers exert. Clearly, surveys require considerable control over the design of data collection tools (i.e., the survey questions and response formats) and the selection of the participating sample. In contrast, semi-structured interviews and focus groups give the researcher control over what task or questions will be posed but much less control over how participants respond or what they substantively contribute. Free response mechanisms are deliberately low-control data collection techniques as the researcher explicitly directs the participants to respond in an individual way (e.g., consider the variety of student responses to a poem in the film *Dead Poets' Society*). Observations of naturally occurring data or inspection of traces left behind by teachers or students only give control to the researcher in terms of which behaviours or objects will be selected for analysis.

Survey Studies. A large number of survey studies about feedback have been conducted, particularly around participant understandings of and attitudes towards feedback and/or their feedback experiences. Effective surveying generally depends on large samples drawn from a clearly defined population. There is greater power to generalise when the sample is fully representative of all key aspects of the population and when the researcher selects participants rather than relying on a convenience sample. The issue of how many participants comprise an appropriate sample size is complex and outside of the scope of this chapter. Interested readers will find some general guidance from NIST/SEMATECH (2003) and Raosoft (2017).

Survey methods are frequently used within the feedback literature. For example, our own work has surveyed students ($N=278$, Brown, Peterson, & Yao, 2016; $N=193$, Harris, Brown, & Harnett, 2014b) and teachers ($N=518$, Brown, Harris, & Harnett, 2012) concerning their understandings of feedback. Indeed, the development of a standardized self-report questionnaire about the function and nature of feedback has been the subject of a series of studies among New Zealand students ($N=499$, 705, 536, respectively) (Brown, Irving, & Peterson, 2009; Irving, Petersen, & Brown, 2007, 2008). Other large to reasonably-large feedback surveys include a study of feedback perceptions among Cantonese-speaking students and teachers ($N=1740$, 460, respectively) (Carless, 2006); a survey of Portuguese university students ($N=502$, in two sites) (Flores, Veiga Simão, Barros, & Pereira, 2015); and a two-study survey of Australian university students (Study 1 $N=57$; Study 2 $N=277$) (Lizzio, & Wilson, 2008). In contrast, smaller scale surveys include Robinson, Hope, and Holyoak (2013) with 166 university students; Jodaie, Farrokhi, and Zoghi (2011) with 100 high school students and 30 teachers; Bayerlein (2014) with 103 undergraduate and 30 postgraduate students; and Bevan, Badge, Cann, Willmott, and Scott (2008) with just 45 students. Clearly, inconsistency in results can arise from the differences in the design of survey questions, response scales, and samples. A constant threat to the validity of claims is the large errors in reported survey results attributable to smaller sample sizes; simply put more is better in survey work (Marsh, Hau, Balla, & Grayson, 1998).

Interviews. An obviously direct way of collecting information concerning people's perspectives, experiences, opinions, attitudes, and beliefs about feedback is to ask them. The interview allows participants to state directly for themselves their own thoughts, ideas, and

emotions. Even quite young interviewees are able to answer questions about intentions, attitudes, thoughts, and experiences. Obviously, interviews that follow a more rigid schedule of questions or topics administered in the same way (de Leeuw, 2008) will generate more comparable information across informants. However, relaxing the structure and sequence of the interview will generate more natural and potentially more insightful responses (Brenner, 2006) and allow interviewers to ask probing questions and follow up on statements of interest. Regardless, institutional review boards or ethics committees will probably require at least a schedule of 'semi-structured' questions to establish that the study will do no harm.

The majority of interview studies about feedback include relatively small numbers of participants. The number of interviewees needed to create a sense of saturation concerning the range of responses is difficult to estimate and the answer seems to depend on many factors, including pragmatic issues of time needed to collect and analyse data (Baker & Edwards, 2012). Nonetheless, recommendations seem to gravitate to at least 20 or 30. Given such small sample sizes are common, how participants are chosen is particularly important. While some sampling strategies (e.g., selecting prospective interviewees from a larger pool based on particular criteria; Harris & Brown, 2016) may help increase the generalizability of findings, caution must be taken when drawing conclusions based on any kind of convenience sample.

There is concern about the potential influence of the interviewer, whose presence and manner may inadvertently influence the informant to say what the person believes is expected by the interviewer, sometimes referred to as socially desirable responding. As interviewer influence can range from mild and inadvertent to oppressive and controlling, researchers must attend carefully to the way knowledge is constructed within an interview context and minimise the potential of interviewer distortion in the interview. The type and nature of the interviewer impact should be acknowledged and considered when interpreting results (Kvale, 2002). The challenge is how to elicit appropriate information without distorting data.

Interview studies commonly focus on the perspectives of teachers (e.g., Bailey & Garner, 2010; Price, Handley, Millar, and O'Donovan, 2010), students (e.g., Bruno and Santos, 2010; Gamlem and Smith, 2013; Higgins, Hartley, and Skelton, 2002; Tunstall & Gipps, 1996a) or both (e.g., Tjeerdsmā, 1997). For example, one study of 48 UK university lecturers showed that the idiosyncratic perceptions and practices of the lecturers did not reflect the intentions of institutional policies (Bailey & Garner, 2010). Because this study was situated in one institution, it is difficult to know if the results generalise beyond this context, though such an expectation seems probable.

An interesting method that potentially reduces interviewer effects is the use of stimulated recall as a trigger for the interview. Recall can be stimulated with samples of previously given or received written feedback or audio or audio-visual recordings of a feedback situation; these real life prompts then become the basis of discussion focused on aspects that seem relevant to the theme of the interview (Dempsey, 2010). Examples of studies that use stimulated recall around feedback experiences are Hargreaves (2013) and Van der Kleij, Adie, and Cumming (2016).

Another useful interview style to consider is the think-aloud. Think-aloud protocols require that individuals conduct a task while verbalising their thoughts, reasons, and reactions while doing a task (Kuusela & Paul, 2000). This information provides insights into the

underlying reasons for why participants do what they do (e.g., Why might teachers provide particular feedback? Why might students respond to feedback in particular ways?). No doubt the challenge with ‘talking about thinking while doing’ is that it depends on the ability of participants to be aware of their thinking and their skill at putting those thoughts into spoken words. Clearly, this requires significant skills and substantial training and prompting to ensure participants do not ‘go quiet’ while they concentrate on a novel task or alternately begin to make inferences about their thinking instead of reporting their actual thoughts (Ericsson & Simon, 1980). At least one think-aloud study has investigated what what students do with the feedback they are given (Handley, Price, & Millar, 2011).

Focus Groups. Increasing the number of participants by creating a focus group can also potentially decrease the influence of the moderator. The focus group method aims to elicit open and rich responses by allowing participants to interact and discuss the topic; because of this group interaction, the discussion is potentially less interviewer-centred (Kitzinger, 1994; Morgan 1997; Rabiee, 2004). Such free-flowing conversations generate data based on group interactions that are different to the sum of individual contributions in an interview setting. Focus group discussions about feedback have included teachers (Irving, Harris, & Peterson, 2011), as well as primary (Burnett & Mandel, 2010), secondary (Peterson & Irving, 2008), and university students (Lipnevich & Smith, 2009a; Poulos & Mahony, 2008). Obviously, the make-up of focus groups matters; pre-existing relationships among group members may make some participants hesitant to share particular opinions, while others may be more open due to friendship. An important element to a well-constructed focus group is having a starter task that ensures all members have something to contribute, reducing the impact of negative social practices, such as ‘loafing’ or ‘free-riding’, collusion, friendship bias, or decibel control of discussion (Pond & Ul-Haq, 1997). For example, an interesting free response technique that has been used as a focus group task is to ask students to draw a picture of their own personal perception of the phenomenon of interest (Harris, Harnett, & Brown, 2009). The fact that the picture is personal and usually not artfully drawn (unless by art students), means it requires explanation and discussion, the very grist of focus groups.

Observation. Gaining access to learning spaces and processes in real-world classroom settings is the antithesis to the laboratory environment of the experiment. Actual behaviours being observed are generally naturalistic once those being observed are accustomed to the presence of the observer or camera. However, researchers have to exercise control of observation by developing sampling systems (i.e., How often should particular behaviours or events be recorded? Whose actions (e.g., the teacher, particular students) are the focus of observation?). Mechanisms for recording data (e.g., checklists; see an example in Voerman, Meijer, Korthagen, & Simons, 2012) have to be developed and checked for validity and reliability. The downside, unsurprisingly, is that the phenomenon of interest may not occur in front of the observer or may not occur very often when it is not elicited by the researcher. This means that observation depends on the visibility and sufficient presence of the behaviour of interest to make the observation of value. Length of observation varies: Voerman et al. (2012) observed only one lesson per teacher to examine feedback practices, while Björklund Boistrup (2010) attended all mathematics lessons for the entire week.

However, even if many observations take place, there is still the potential of researcher impact upon results (e.g., the potential that Margaret Mead was deceived by her informants;

Kawulich, 2005). Observation studies of classroom feedback processes often are augmented by complementary methods (e.g., interviews or focus groups) to ensure that the observed behaviours are interpreted from subject rather than researcher perspectives (e.g., Harris & Brown, 2013). Observed data is often a powerful starting point for stimulated recall interviews, such as the ones discussed in the previous section.

Traces. A related data collection technique to observation is the study of the traces or artefacts teachers and students create as a result of their normal classroom practice (e.g., comments left on essays or course books). The collection and analysis of traces left behind in natural processes is an old and well-established tradition in social science research (Webb, Campbell, Schwartz, Sechrest, & Grove, 1981). These traces are created naturally without any input from the researcher and so have great validity, although sampling bias can occur (e.g., when participants only select what they consider to be their ‘best’ examples to share). Manual trace data dominate research into feedback currently, with much arising from written teacher feedback comments (Hyland, & Hyland, 2001; Lee, 2007; Matsumura, Patthey-Chavez, Valdes, & Garnier, 2002; Ruiz-Primo, & Li, 2013). Similar data can come from student written comments to their peers or themselves as part of peer or self-assessment practices (Harris, Brown, & Harnett, 2014a).

Research which analyses examples of feedback online has also gained popularity, made possible by increased computer and social media usage alongside the development of digital tools which allow researchers to have access to significant information about the person’s online behaviours (e.g., time spent, places visited, and so on). Two examples of digital log files being used as a way of exploring student feedback use can be seen in Narciss et al.’s (2014) use of log-files recorded in a web-based intelligent learning environment to trace student actions on a task following varying types of feedback, and Harper, Brown, and Hill’s (2016) use of log-files of students’ voluntary use of hints in an online teaching and testing resource. These digital traces are the core material of ‘big data’ research (Hofferth et al., 2017) and may be a potential source for further research.

Case Studies. While studying a large sample of individuals or sites has potential to create robust estimates of parameters, it is sometimes useful to examine in great detail how things take place in a much narrower sample of situations. Researchers often choose to study smaller samples much more intensively with multiple methods and modes of data collection. Such rich but small-scale investigations, often called case studies, may be particularly valuable to feedback research because the complexities of how feedback works are still being clarified. Case studies involve purposefully selecting a small sample of individuals or sites that are studied intensively, often making use of multiple methods so as to create a deep understanding of the complexity of factors acting upon a phenomenon (Yin, 2006).

For example, Harris and Brown (2013) selected three teachers out of a pool of 26 interviewees and intensively studied their classroom assessment and feedback practices. Data collection involved video-recorded observations, teacher interviews, and student focus groups. Other examples of case study include: Bruno and Santos (2010), who studied three middle school students’ experiences of written feedback via interviews with the students, observation of them in their classroom environment, and analyses of their assignments; Hyland (2003), who studied six students via interviews and analysis of teacher feedback on their written work; and Li and Barnard (2011), who studied the feedback beliefs and practices

of a group of tutors via survey, individual interviews, ‘think aloud’ and stimulated recall sessions, and focus group interviews. This approach, especially when combined with multiple observations or data types, can create powerful and insightful descriptions of how processes in a domain take place.

Mixing Methods

Throughout this grand tour of data collection methods, many studies use more than one method. This is popularly known as mixed methods (Johnson & Onwuegbuzie, 2004), because methods that have different assumptions about knowledge are brought together to investigate a phenomenon. However, it may be better to think of this approach as ‘mixing’ methods (Brannen, 1992) since each method is integral in itself and, as yet, there do not appear to be methods that somehow inherently cross paradigmatic boundaries. The case studies described above mix methods; likewise, several studies have used two or more methods sequentially to understand feedback. For example, Lizzio and Wilson (2008) used content analysis of student comments to create a construct frame for a subsequent survey study.

Mixing methods has a long provenance in multi-method research (Brewer & Hunter, 1989), where multiple methods, usually drawn from a pool of philosophically aligned methods (e.g., surveys with multiple batteries), are used to investigate a phenomenon. More recently, mixing of methods has focused on crossing paradigmatic boundaries so as to increase the probability of fully understanding a phenomenon by approaching it from multiple perspectives, multiple informants, and multiple data sources. The logic of triangulation is invoked as a rationale for mixing methods, though considerable caution needs to be exercised in believing that contrasting methods will deliver corroboration (Harris & Brown, 2010; Smith, 2006). Nonetheless, increasingly feedback studies are using multiple or mixed methods to overcome method artefacts.

Data analysis

Having collected the data, the researcher must turn that raw information into interpretation using systematic, rigorous methods of analysis. Generally, the goal is to document procedures such that readers can verify the credibility of subsequent interpretation. Data analytic techniques group into two major classes: (a) content classification and (b) statistical analysis. While statistical techniques of analysis may be governed by conventions and rules, there is still a subjective element in deciding which technique to apply and which standards to adopt (Brown, 2016). Hence, a sufficient warrant for findings depends upon credible techniques of analysis.

Content classification methods

A variety of classification methods exist. All these methods require analysts to assign a categorical attribution to elements of the raw data (Bartholomew, Henderson, & Marcia, 2000). Rules or rubrics for assigning data objects to classifications can be derived from a priori theoretical categories (i.e., deductive analysis) or derived from a close inspection of the data (i.e., inductive analysis).

Deductive analysis of data involves mapping aspects of the raw data to pre-existing categories derived from a theory or model of how the phenomenon should work (Elo & Kyngäs, 2008). A large body of inductive methods exist (e.g., grounded theory, phenomenography, content analysis, thematic analysis, etc.), but there are many similar

features to such methods (Thomas, 2006). General principles involve discovering thematic categories that efficiently and effectively group common phenomena within the data and which allow for a simplification of the data (Braun & Clarke, 2006). Validation of classification can be determined through having two or more analysts conduct coding of a sufficient sample to establish a measure of consensus between analysts (e.g., % agreement) or consistency of classification (e.g., Kappa or Krippendorff coefficients) (Stemler, 2004).

Deductive analysis. A number of studies have adopted pre-existing feedback frameworks and used them to guide analysis of feedback data. For example, Hattie and Timperley's (2007) four levels of feedback framework (i.e., self, task, process, and self-regulation) has been used to classify the types of feedback students give each other (Harris, Brown, & Harnett, 2014a) and which teachers give students (Harris, Harnett, & Brown, 2013). Other studies that used the Hattie and Timperley framework to guide analysis include Ajjawi and Boud (2017), Björklund Boistrup (2010), and Gan and Hattie (2014). Another framework for coding feedback was developed by Tunstall and Gipps (1996b), and this has also been used by many other authors (e.g., Parr and Timperley, 2010). These are just two of the frameworks available for other researches to draw upon, with a further example being Chi's (1996) four types of feedback: Corrective, Reinforcing, Didactic, and Suggestive (Tseng & Tsai, 2007).

Inductive analysis. As mentioned already, a large number of inductive techniques have been used to investigate feedback. The rationale for this choice of technique is not always obvious and, unsurprisingly, results are extremely variable since quite different techniques are used. An example of a reasonably replicable technique arising from the research goals is seen in Van der Kleij et al. (2016) who, among other things, measured the length of each feedback conversation and counted the number of words spoken by teacher and student, along with the number of times each person stopped the replayed video recording to comment on what was happening in the pre-recorded student-teacher feedback conversation. More difficult to replicate are such techniques as: content analysis (e.g., Mayring, 2000; Weaver, 2006), as used by Lizzio and Wilson (2008); phenomenology (e.g., Groenewald, 2004), as used by Orsmond and Merry (2011); or phenomenography (e.g., Marton, 1981; Harris, 2011), as used by McLean, Bond, and Nicholson (2014). The adoption of these techniques can allow novel and important discoveries about not just categories of interest but also the relationship of such phenomena to each other. The challenge facing inductive analysis is to establish credibility in the process of classification and the potential replicability of the classification systems. Some inductive analytic studies in feedback report the development of robust coding schema and the testing of the schema through the ability of independent judges to reasonably replicate each other's classification (e.g., Lizzio & Wilson, 2008; Nelson & Schunn, 2009; Parr & Timperley, 2010), but many could be improved via more rigorous reporting of the coding process.

Statistical analytic methods

Survey and experimental studies about feedback tend to be heavy users of statistics to make sense of results. A major goal in statistical analysis is to identify whether the observed results are unlikely to be due to chance (null hypothesis significance testing) and how large and substantive the results are (i.e., effect size) (Field, 2016). A common early step in many statistical studies is the evaluation and determination of measurement quality, which is a substantial challenge when attempting to grapple with psychological phenomena (Crocker,

2006). The evaluation of scale attributes assumes that multiple indicators are needed to reduce error in measuring complex phenomena and is often summarised, albeit inappropriately (Sijtsma, 2009; Teo & Fan, 2013), through Cronbach's alpha estimate of internal reliability, and more robustly through model to data fit indices generated by confirmatory factor analysis (Bandalos & Finney, 2010). Once robust measurement properties are established (e.g., scale reliability, fit to the data, validation evidence for the constructs being measured), many studies go further into determining whether paths between constructs are both statistically and practically significant. While conventional regression analysis can achieve this, structural equation modeling is a more powerful causal-correlational analysis approach since it retains latent structures and errors and allows for correlated predictors and outcomes (Hoyle, 1995). The causal paths introduced in such analysis, of course, are not proven in non-experimental studies. Alternately, rather than establish regressions among constructs, analysis may examine whether there are differences in responses (e.g., mean scores or parameter estimates) between participant groups (i.e., analysis of variance and invariance testing, respectively). The use of such techniques can test empirically the credibility of theoretical claims about feedback and generate hypotheses about the relationship of feedback factors to outcomes.

Large scale survey studies that have used a combination of factor analytic techniques (to establish measurement properties) and structural equation modeling (to establish potentially causal relationships among constructs) include: Brown, Harris, & Harnett, (2012); Brown, Peterson, and Yao (2016), Burnett (2002); Harks et al. (2014); Harris, Brown, and Harnett (2014b); Rakoczy et al. (2013). A simpler approach than structural equation modeling is to correlate factor scores once these are determined (e.g., Lizzio & Wilson, 2008). This approach reduces the latent score to a single scale score and does not establish directionality in paths. Correlations, which are non-directional linear relationships, can be used to identify relationships, without suggesting causal direction (Nelson & Schunn, 2009; Parr & Timperley, 2010).

An even simpler approach is to determine a set of factors or scales and generate mean scale scores for participants to examine differences between time, group, or condition. Reporting the means and standard deviations can permit eyeball inspection as to whether the means differ (Tang & Harrison, 2011), but a more sophisticated approach is to use an inferential technique such as analysis of variance (Nesbit & Burton, 2006) and multivariate analysis of variance (Kim, Lee, Chung, & Bong, 2010; Kerssen-Griep, & Witt, 2012) to determine not only whether the observed differences are unlikely to occur by chance but also to establish the size of difference after taking into account the variability of scores within each group. When non-continuous variables are used (e.g., proportion of feedback falling into certain categories), statistical methods exist (e.g., McNemar test) to determine whether there are differences in the distribution of these categorisations between times or groups (e.g., Parr & Timperley, 2010).

Conclusion

This examination of methods used in feedback studies may read somewhat like a research methods textbook, but the concerns readers should have about research results often have to do with researchers ignoring well-established threats to the validity of their claims due to the methodological choices made. What this means is that editors and reviewers of studies need

to insist that authors provide enough detail about methodological choices to permit critical evaluation and potential replication. While statistical analytic studies are potentially easy to critique (e.g., if a parametric test is applied to non-parametric data), it is somewhat more challenging to critique subjective coding if authors simply report that “themes emerged from the data”. The importance of clarity and completeness in reporting of methods decisions cannot be over-emphasised. It also seems clear from this review that the method issues raised in the general literature apply equally to research into feedback.

An important aspect of research into feedback that we have not seen extensively in the field is attention to the importance and influence of cultural context upon results. It is highly likely that there are differences in how various groups respond to methods that impact upon results (e.g., starting values for Asian groups may be more modest than Western groups on items having to do with self-evaluations of behaviours), yet very few studies include participants from different countries (an exception is Furneaux, Paran, & Fairfax, 2007). Hence, the comparability of data collection methods across contexts may be much more problematic than expected. Translation and adaptation of instruments across languages and contexts is not guaranteed by ‘translation, back-translation’ procedures since the meaning of a phenomenon in a different context is not necessarily captured by literal rather than functional equivalence logic (Werner & Campbell, 1973). Our own study with a self-report survey inventory of teacher conceptions of feedback generated completely different analytic structures and incompatible models between New Zealand and Louisiana, United States which was attributed to the very different policy frameworks around assessment and feedback (Brown, Harris, O’Quin, & Lane, 2017). This means that identical replication is probably an extremely laudable but potentially impossible goal when attempting to cross large cultural, social, or policy boundaries. This suggests that there will always be discrepancy and inconsistency in results; it will always probably ‘depend’. This issue means that researcher users must be careful about how research findings are applied in a different context.

Nevertheless, researchers have to make decisions about ‘best’ or most appropriate method. Going back to the big questions of what we need to know about feedback in this context is perhaps the most defensible way to approach feedback research. It is outside the scope of this chapter to decide what those questions are, but we encourage researchers to consider: What are the big questions we need to answer about feedback? What kinds of methodologies or designs might be appropriate to investigate these questions? What new or innovative approaches might be worth adopting? Considering questions around ‘what we know we don’t know’ would be a good start. For example, we suggest that the field needs a deeper understanding of what students actually do or do not do with feedback and why those choices are made. The teaching and curriculum world is creating new modes of feedback delivery (e.g., computerised feedback, feedback from peers/self) and how these can be structured to ensure positive impact on learning is still uncertain. The one fMRI study suggests that potential insights into the effect of feedback can be found using this high technology approach, though the validity of such studies is still in question (Vul, Harris, Winkielman, & Pashler, 2009).

It is clear in reading the literature that feedback elicits highly personal responses, partly because feedback is so complex (Lipnevich, Berg, & Smith, 2016). Having a clear understanding of the participants in a study is important if we hope to make any claims about

what is or is not effective. Participants (i.e., students, teachers, etc.) are not homogenous groups; hence, methods need to be able identify and account for within group variation.

Our conclusion is that the world of feedback research has a long way to go to establish robust and consistent findings due, in part, to the lack of clarity in reporting how research was conducted and the lack of consistency in methods, samples, and tools across studies. The field is being enriched by many small and exploratory studies, but many claims about feedback are not well and consistently supported by data. It is necessary that educational decisions about feedback are informed by approaches which are empirically supported rather than ideologically driven. It is hoped this review will contribute to the development of a more robustly described and conducted body of research into feedback. We encourage increased scrutiny from research users, which will lead to more evidence based feedback strategies being implemented in schools for the benefit of learners.

References

- Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment & Evaluation in Higher Education*, 42(2), 252-265. doi:10.1080/02602938.2015.1102863
- Ashwell, T. (2000). Patterns of Teacher Response to Student Writing in a Multiple-Draft Composition Classroom: Is Content Feedback Followed by Form Feedback the Best Method? *Journal of Second Language Writing*, 9(3), 227-257. doi:10.1016/S1060-3743(00)00027-8
- Autoimmunity Research Foundation. (2012). Differences between in vitro, in vivo, and in silico studies. Retrieved 12 November, 2015, from http://mpkb.org/home/patients/assessing_literature/in_vitro_studies
- Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education*, 15(2), 187-198. doi:10.1080/13562511003620019
- Baker, S. E., & Edwards, R. (2012). *How many qualitative interviews is enough? Expert voices and early career reflections on sampling and cases in qualitative research*. Southampton, UK: National Centre for Research Methods. Retrieved from http://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York: Routledge.
- Bartholomew, K., Henderson, A. J. Z., & Marcia, J. E. (2000). Coded semistructured interviews in social psychological research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 286-312). Cambridge, UK: Cambridge University Press.
- Bayerlein, L. (2014). Students' feedback preferences: how do students react to timely and automatically generated assessment feedback? *Assessment & Evaluation in Higher Education*, 39(8), 916-931. doi:10.1080/02602938.2013.870531
- Bevan, R., Badge, J., Cann, A., Willmott, C., & Scott, J. (2008). Seeing Eye-to-Eye? Staff and Student Views on Feedback. *Bioscience Education*, 12(1), 1-15. doi:10.3108/beej.12.1

- Björklund Boistrup, L. (2010). *Assessment discourse in mathematics classrooms: A multimodal social semiotic study*. PhD, Stockholm University, Stockholm.
- Brannen, J. (Ed.). (1992). *Mixing methods: Qualitative and quantitative research*. Aldershot, UK: Avebury Ashgate Publishing.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. 10.1191/1478088706qp063oa
- Brenner, M. E. (2006). Interviewing in educational research. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 357-370). Mahwah, NJ: Lawrence Erlbaum Associates.
- Brewer, J., & Hunter, A. (1989). *Multimethod Research: A Synthesis of Styles*. Newbury Park, CA: SAGE.
- Brown, G. T. L. (2016). The qualitative secret within quantitative research: It's not just about numbers. In C. J. McDermott & Kožuh, B. (Eds.). *Modern Approaches in Social and Educational Research* (pp. 33-42). Antioch University: Los Angeles, CA.
- Brown, G. T. L., Harris, L. R., & Harnett, J. (2012). Teacher beliefs about feedback within an Assessment for Learning environment: Endorsement of improved learning over student well-being. *Teaching and Teacher Education*, 28(7), 968-978.
doi:10.1016/j.tate.2012.05.003
- Brown, G. T. L., Harris, L. R., O'Quin, C., & Lane, K. E. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: identifying and understanding non-invariance. *International Journal of Research & Method in Education*, 40(1), 66-90. 10.1080/1743727X.2015.1070823
- Brown, G. T. L., Irving, S. E., & Peterson, E. R. (2009, August). *The more I enjoy it the less I achieve: The negative impact of socio-emotional purposes of assessment and feedback on academic performance*. Paper presented at biennial conference of the European Association for Research in Learning and Instruction, Amsterdam, NL.
- Brown, G. T. L., Peterson, E. R., & Yao, E. (2016). Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement. *British Journal of Educational Psychology*, 86(4), 606-629. doi:10.1111/bjep.12126
- Bruno, I., & Santos, L. (2010). Written Comments as a Form of Feedback. *Studies in Educational Evaluation*, 36(3), 111-120.
- Burnett, P. C. (2002). Teacher Praise and Feedback and Students' Perceptions of the Classroom Environment. *Educational Psychology*, 22(1), 5-16.
doi:10.1080/01443410120101215
- Burnett, P. C., & Mandel, V. (2010). Praise and Feedback in the Primary Classroom: Teachers' and Students' Perspectives. *Australian Journal of Educational & Developmental Psychology*, 10, 145-154.
- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290-298.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233. doi:10.1080/03075070600572132
- Chase, J. A., & Houtmanfar, R. (2009). The differential effects of elaborate feedback and basic feedback on student performance in a modified, personalized system of

- instruction course. *Journal of Behavioral Education*, 18(3), 245-265.
doi:10.1007/s10864-009-9089-2
- Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, 10, 33–49.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction*, 20(4), 328-338.
- Corbalan, G., Kester, L., & J.G. van Merriënboer, J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, 19(6), 455-465.
- Crocker, L. (2006). Introduction to measurement theory. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of Complementary Methods in Education Research* (pp. 371-384). Mahwah, NJ: Lawrence Erlbaum Associates.
- de Leeuw, E. (2008). Self-administered questionnaires and standardized interviews. In P. Alasuutari, L. Bickman & J. Brannen (Eds.), *The SAGE handbook of social research methods* (pp. 313-327). London: SAGE.
- Dempsey, N. P. (2010). Stimulated Recall Interviews in Ethnography. *Qualitative Sociology*, 33(3), 349-367. doi:10.1007/s11133-010-9157-x
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6(621) 10.3389/fpsyg.2015.00621
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, 62(1), 107–115. doi: 10.1111/j.1365-2648.2007.04569.x
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, 87(3), 215-251. 10.1037/0033-295X.87.3.215
- Field, A. (2016). *An Adventure in Statistics: The Reality Enigma*. London: Sage.
- Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education*, 40(9), 1523-1534.
doi:10.1080/03075079.2014.881348
- Furneaux, C., Paran, A., & Fairfax, B. (2007). Teacher stance as reflected in feedback on student writing: An empirical study of secondary school teachers in five countries. *IRAL - International Review of Applied Linguistics in Language Teaching*, 45(1), 69-94. doi: 10.1515/iral.2007.003
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150-169.
doi:10.1080/0969594x.2012.749212
- Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, 42(6), 861-878. 10.1007/s11251-014-9319-4
- Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4), 304-315.
- Groenewald, T. (2004). A Phenomenological Research Design Illustrated. *International Journal of Qualitative Methods*, 3(1), 1-26.
- Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40–53.
doi:10.1016/j.jslw.2007.01.001

- Hamer, J., Purchase, H., Luxton-Reilly, A., & Denny, P. (2015). A comparison of peer and tutor feedback. *Assessment & Evaluation in Higher Education*, 40(1), 151-164. doi:10.1080/02602938.2014.893418
- Handley, K., Price, M., & Millar, J. (2011). Beyond 'doing time': investigating the concept of student engagement with feedback. *Oxford Review of Education*, 37(4), 543-560. doi:10.1080/03054985.2011.604951
- Harber, K. D., Gorman, J. L., Gengaro, F. P., Butisingh, S., Tsang, W., & Ouellete, R. (2012). Students' race and teachers' social support affect the positive feedback bias in public schools. *Journal of Educational Psychology*, 104(4), 1149-116.
- Hargreaves, E. (2013). Inquiring into children's experiences of teacher feedback: Reconceptualising Assessment for Learning. *Oxford Review of Education*, 39(2), 229-246. doi:10.1080/03054985.2013.787922
- Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: the role of feedback's perceived usefulness. *Educational Psychology*, 34(3), 269-290. doi:10.1080/01443410.2013.785384
- Harper, A., Brown, G. T. L., & Hill, M. F. (2016, July). *Biology students' use of online feedback in formative assessment*. Poster presented at the biennial conference of the International Test Commission, Vancouver, Canada.
- Harris, L. R. (2011). Phenomenographic perspectives on the structure of conceptions: The origins, purposes, strengths, and limitations of the what/how and referential/structural frameworks. *Educational Research Review*, 6(2), 109-124. 10.1016/j.edurev.2011.01.002
- Harris, L. R., & Brown, G. T. L. (2010). Mixing interview and questionnaire methods: Practical problems in aligning data. *Practical Assessment Research & Evaluation*, 15(1), Available online: <http://pareonline.net/pdf/v15n11.pdf>.
- Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111. doi:10.1016/j.tate.2013.07.008
- Harris, L. R., & Brown, G. T. L. (2016, April). "Everything you do is giving them feedback": A phenomenographic study of teacher conceptions of feedback. Paper presented at the American Educational Research Association Annual Meeting, Washington DC.
- Harris, L. R., Harnett, J. A., & Brown, G. T. L. (2009). 'Drawing' out student conceptions: Using pupils' pictures to examine their conceptions of assessment. In D. M. McInerney, G. T. L. Brown & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning*. (pp. 321-330). Charlotte, NC US: Information Age Publishing.
- Harris, L. R., Brown, G. T. L., & Harnett, J. (2014a). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: Applying Hattie & Timperley's feedback model. *Assessment in Education: Principles, Policy and Practice*, 22(2), 265-281. doi:10.1080/0969594X.2014.976541
- Harris, L. R., Brown, G. T. L., & Harnett, J. (2014b). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional

- responses. *Educational Assessment, Evaluation and Accountability*, 26(2), 107-133. doi:10.1007/s11092-013-9187-5
- Harris, L. R., Harnett, J., & Brown, G. T. L. (2013, April). *Exploring the content of teachers' feedback: What are teachers actually providing to students?* Paper presented at the annual AERA conference, San Francisco, CA.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Higgins, R., Hartley, P., & Skelton, A. (2002). The Conscientious Consumer: Reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, 27(1), 53-64. doi:10.1080/03075070120099368
- Hofferth, S. L., Moran, E. F., Entwistle, B., Aber, J. L., Brady, H. E., Conley, D., . . . Hubacek, K. (2017). Introduction: History and Motivation. *The ANNALS of the American Academy of Political and Social Science*, 669(1), 6-17. doi:10.1177/0002716216682715
- Hoyle, R. H. (1995). The structural equation modeling approach: Basic concepts and fundamental issues. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 1-15). Thousand Oaks, CA: Sage.
- Hunter, J. E. (2001). The Desperate Need for Replications. *Journal of Consumer Research*, 28(1), 149-158. doi:10.1086/321953
- Hyland, F. (2003). Focusing on form: student engagement with teacher feedback. *System*, 31(2), 217-230. doi:10.1016/s0346-251x(03)00021-6
- Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10(3), 185-212. doi:10.1016/s1060-3743(01)00038-8
- Irving, S. E., Harris, L. R., & Peterson, E. R. (2011). "One assessment doesn't serve all the purposes" or does it?: New Zealand teachers describe assessment and feedback. *Asia-Pacific Education Review*, 12(3), 413-426. doi:10.1007/s12564-011-9145-1
- Irving, S. E., Petersen, E. R., & Brown, G. T. L. (2007, August). *Student conceptions of feedback: A study of New Zealand secondary students*. Paper presented at the Biannual Conference of the European Association for Research in Learning and Instruction (EARLI), Budapest, Hungary.
- Irving, S. E., Peterson, E. R., & Brown, G. T. L. (2008, July). *Feedback and academic achievement: The relationship between students' conceptions of feedback and achievement*. Paper presented at the 6th Biennial Conference of the International Test Commission, Liverpool, UK
- Jodaie, M., Farrokhi, F., & Zoghi, M. (2011). A Comparative Study of EFL Teachers' and Intermediate High School Students' Perceptions of Written Corrective Feedback on Grammatical Errors. [Article]. *English Language Teaching*, 4(4), 36-48. doi:10.5539/elt.v4n4p36
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
- Kawulich, B. B. (2005). Participant Observation as a Data Collection Method. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), <http://www.qualitative-research.net/index.php/fqs/article/view/466/996>.

- Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, 27(5), 411-425.
- Kerssen-Griep, J., & Witt, P. L. (2012). Instructional Feedback II: How do instructor immediacy cues and facework tactics interact to predict student motivation and fairness perceptions? *Communication Studies*, 63(4), 498-517.
doi:10.1080/10510974.2011.632660
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Kim, S.-I., Lee, M.-J., Chung, Y., & Bong, M. (2010). Comparison of brain activation during norm-referenced versus criterion-referenced feedback: The role of perceived competence and performance-approach goals. *Contemporary Educational Psychology*, 35(2), 141-152.
- Kitzinger, J. (1994). The methodology of Focus Groups: the importance of interaction between research participants. *Sociology of Health & Illness*, 16(1), 103-121.
10.1111/1467-9566.ep11347023
- Kuusela, H., & Paul, P. (2000). A Comparison of Concurrent and Retrospective Verbal Protocol Analysis. *The American Journal of Psychology*, 113(3), 387-404.
10.2307/1423365
- Kvale, S. (2002, March 7-9). *Dialogue as Oppression and Interview Research*. Paper presented at the Nordic Educational Research Association Conference Tallinn, Estonia. Retrieved from <http://bit.ly/2mROjhH>
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing*, 12(3), 180-198.
- Li, J., & Barnard, R. (2011). Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study. *Assessing Writing*, 16(2), 137-148. doi:10.1016/j.asw.2011.02.004
- Lindsay, D. S. (2015). Replication in Psychological Science. *Psychological Science*, 26(12), 1827-1832. doi:10.1177/0956797615616374
- Lipnevich, A. A., & Smith, J. (2009a). "I really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation & Accountability*, 21(4), 347-367. doi:10.1007/s11092-009-9082-2
- Lipnevich, A. A., & Smith, J. K. (2009b). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319-333. doi:10.1037/a0017841
- Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown & L. R. Harris (Eds.), *The Handbook of Human and Social Conditions in Assessment* (pp. 169-185). New York: Routledge.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263-275.
doi:10.1080/02602930701292548
- Maier, U. (2010). Accountability policies and teachers' acceptance and usage of school performance feedback - a comparative study. *School Effectiveness & School Improvement*, 21(2), 145-165. doi:10.1080/09243450903354913

- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research. *Perspectives on Psychological Science*, 7(6), 537-542. doi:10.1177/1745691612460688
- Marsh, H. W., Hau, K.-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181-220. 10.1207/s15327906mbr3302_1
- Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science*, 10, 177-220.
- Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher Feedback, Writing Assignment Quality, and Third-Grade Students' Revision in Lower- and Higher-Achieving Urban Schools. *Elementary School Journal*, 103(1), 3-25.
- Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2), <http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385>.
- McLean, A. J., Bond, C. H., & Nicholson, H. D. (2015). An anatomy of feedback: a phenomenographic investigation of undergraduate students' conceptions of feedback. *Studies in Higher Education*, 40(5), 921-932. 10.1080/03075079.2013.855718
- Morgan, D. L. (1997). *Focus groups as qualitative research* (2nd ed.). Thousand Oaks, CA: SAGE.
- Narciss, S., Sosnovsky, S., Schnaubert, L., Andrès, E., Eichelmann, A., Goguadze, G., & Melis, E. (2014). Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Computers & Education*, 71(0), 56-76. doi:10.1016/j.compedu.2013.09.011
- National Research Council. (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375-401. doi:10.1007/s11251-008-9053-x
- Nesbit, P. L., & Burton, S. (2006). Student justice perceptions following assignment feedback. *Assessment & Evaluation in Higher Education*, 31(6), 655-670.
- Neuliep, J. W. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, 5(4), 85.
- NIST/SEMATECH. (2003). *e-Handbook of Statistical Methods*. Retrieved from <http://www.itl.nist.gov/div898/handbook>
- Orsmond, P., & Merry, S. (2011). Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, 36(2), 125-136. doi:10.1080/02602930903201651
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15, 68-85. doi:10.1016/j.asw.2010.05.004
- Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, 29(0), 115-124. doi:10.1016/j.learninstruc.2013.09.002
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91-105. doi:10.1207/S15326985EP3702_4

- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238-250. doi:10.1016/j.learninstruc.2007.05.001
- Pond, K., & Ul-Haq, R. (1997). Learning to assess students using peer review. *Studies in Educational Evaluation*, 23(4), 331-348. doi:10.1016/S0191-491X(97)86214-1
- Poulos, A., & Mahony, M. J. (2008). Effectiveness of feedback: the students' perspective. *Assessment & Evaluation in Higher Education*, 33(2), 143 - 154. doi:10.1080/02602930601127869
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277-289. doi:10.1080/02602930903541007
- Rabiee, F. (2004). Focus group interview and data analysis. *Proceedings of the Nutrition Society*, 63(4), 655–660. 10.1079/PNS2004399
- Rakoczy, K., Harks, B., Klieme, E., Blum, W., & Hochweber, J. (2013). Written feedback in mathematics: Mediated by students' perception, moderated by goal orientation. *Learning & Instruction*, 27, 63-73. doi:10.1016/j.learninstruc.2013.03.002
- Raosoft. (2017). Sample size calculator. Retrieved 28 March, 2017, from <http://www.raosoft.com/samplesize.html>
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415. doi:10.1080/02602930500099193
- Robinson, S., Pope, D., & Holyoak, L. (2013). Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment & Evaluation in Higher Education*, 38(3), 260-272. doi:10.1080/02602938.2011.629291
- Ruiz-Primo, M. A., & Li, M. (2013). Analyzing Teachers' Feedback Practices in Response to Students' Work in Science Classrooms. *Applied Measurement in Education*, 26(3), 163-175. doi:10.1080/08957347.2013.793188
- Schildkamp, K., & Visscher, A. (2010). The use of performance feedback in school improvement in Louisiana. *Teaching and Teacher Education*, 26, 1389-1403.
- Sendziuk, P. (2010). Sink or Swim? Improving Student Learning through Feedback and Self-Assessment. *International Journal of Teaching and Learning in Higher Education*, 22(3), 320-330.
- Shadish, W. R., & Luellen, J. K. (2006). Quasi-experimental design. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 539-550). Mahwah, NJ: LEA.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. doi:10.1007/S11336-008-9101-0
- Simons, D. J. (2014). The Value of Direct Replication. *Perspectives on Psychological Science*, 9(1), 76-80. doi:10.1177/1745691613514755
- Smith, M. L. (2006). Multiple methodology in education research. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 457-475). Mahwah, NJ: LEA.
- Smith, J. K., Smith, L. F., & Smith, B. K. (2017). The reproducibility crisis in psychology: Attack of the clones or phantom menace? . In M. C. Makel & J. A. Plucker (Eds.), *Toward a more perfect psychology: Improving trust, accuracy, and transparency in research* (pp. 273-287). Washington, DC: American Psychological Association.

- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), Available online: <http://tinyurl.com/m7grwph>.
- Strijbos, J.-W., Narciss, S., & Dünnebier, K. (2010). Peer feedback content and sender's competence level in academic writing revision tasks: Are they critical for feedback perceptions and efficiency? *Learning and Instruction*, 20(4), 291-303.
- Tang, J., & Harrison, C. (2011). Investigating university tutor perceptions of assessment feedback: three types of tutor beliefs. *Assessment & Evaluation in Higher Education*, 36(5), 583-604. doi:10.1080/02602931003632340 (uses FA)
- Teo, T., & Fan, X. (2013). Coefficient Alpha and beyond: Issues and alternatives for educational research. *Asia-Pacific Education Researcher*, 22(2), 209-213. 10.1007/s40299-013-0075-z
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237-246. 10.1177/1098214005283748
- Tjeerdsma, B. L. (1997). Comparison of teacher and student perspectives of tasks and feedback. *Journal of Teaching in Physical Education*, 16(4), 388-400. doi:10.1123/jtpe.16.4.388
- Tseng, S.-C., & Tsai, C.-C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4), 1161-1174. doi:10.1016/j.compedu.2006.01.007
- Tunstall, P., & Gipps, C. (1996a). 'How does your teacher help you to make your work better?' Children's understanding of formative assessment. *Curriculum Journal*, 7(2), 185-203.
- Tunstall, P., & Gipps, C. (1996b). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389-404.
- Van der Kleij, F., Adie, L., & Cumming, J. (2016). Using video technology to enable student voice in assessment feedback. *British Journal of Educational Technology*, Advanced online publication. doi:10.1111/bjet.12536
- Voerman, L., Meijer, P. C., Korthagen, F. A. J., & Simons, R. J. (2012). Types and frequencies of feedback interventions in classroom interaction in secondary education. *Teaching and Teacher Education*, 28(8), 1107-1115. doi:10.1016/j.tate.2012.06.006
- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15(6), 589-602.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspectives on Psychological Science*, 4(3), 274-290. doi:10.1111/j.1745-6924.2009.01125.x
- Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379-394. doi:10.1080/02602930500353061
- Webb, E. J., Campbell, D. T., Schwartz, R. D., Sechrest, L., & Grove, J. B. (1981). *Nonreactive Measures in the Social Sciences* (2nd ed.). Boston, MA: Houghton Mifflin.
- Werner, O., & Campbell, D. T. (1973). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A Handbook of Method in Cultural Anthropology* (pp. 398-420). New York: Columbia University Press.

- Yin, R. K. (2006). Case study methods. In J. Green, G. Camilli & P. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 111-122). Mahwah, New Jersey: Lawrence Erlbaum Associates Inc.
- Zumbo, B. D. (2015). *Consequences, side effects and the ecology of testing: Keys to considering assessment in Vivo*. Paper presented at the Plenary address to the 2015 annual conference of the Association for Educational Assessment—Europe (AEA-E), Glasgow, Scotland.