



<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in . This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Sterling, S., Plonsky, L., Larsson, T., Kytö, M., Yaw, K. (2023)

Introducing and illustrating the Delphi method for applied linguistics research

*Research Methods in Applied Linguistics*, 2(1): 100040

<https://doi.org/10.1016/j.rmal.2022.100040>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-489739>

# Introducing and Illustrating the Delphi Method for Applied Linguistics Research

Scott Sterling,<sup>1</sup> Luke Plonsky,<sup>2</sup> Tove Larsson,<sup>2,3</sup> Merja Kytö,<sup>3</sup> & Katherine Yaw<sup>4</sup>

<sup>1</sup>Indiana State University;

<sup>2</sup>Northern Arizona University;

<sup>3</sup>Uppsala University;

<sup>4</sup>University of South Florida

## Abstract

This paper briefly explains how to conduct a Delphi study, an interactive and iterative data collection method that utilizes expert opinion to create an artifact. Using phase one of a larger ongoing project as an example case, this paper explores decision points of the project and discusses limitations and considerations that often go undisclosed in published manuscripts using the Delphi method. When applying the method in our project, we collected four rounds of feedback from an expert panel to create a taxonomy of questionable research practices (QRPs). While our project did not encounter major issues, we did face challenges, for example, sampling, timing, and power dynamics. Readers of this paper will gain an understanding of the basics of conducting a Delphi study along with a better understanding of the behind-the-scenes decisions that many scholars make when conducting research of this kind.

**Keywords:** Delphi method, iterative data collection, questionable research practices, research ethics, expert consensus

## Introduction

### *Contextualizing Our Original Project*

Recent research in applied linguistics has found high rates of Questionable Research Practices (QRPs) and unethical work (Isbell et al., 2021), indicating a need to research these actions in more detail. QRPs are the murky waters of research ethics that fall between ideal behavior and absolute misconduct such as fabrication, falsification, and plagiarism (Steneck, 2006). Similarly, the U.S.'s National Academies of Science defines QRPs as “actions that violate traditional values of the research enterprise and that may be detrimental to the research process”, but that do not “directly damage the integrity of the research process” (Committee on Science Engineering and Public Policy, 1992). Applied linguistics is a field that brings together scholars from a variety of research backgrounds who use a host of different approaches to collect and analyze data. Historically, QRP and research ethics education has been limited in the field (Sterling & Gass, 2017) though we have

seen a steady rise in the number of manuscripts produced in recent years (Yaw, Plonsky, Larsson, Sterling, & Kytö, under review). Other fields such as education, medicine, anthropology, and psychology have a more robust history of considering issues related to research ethics and appear to more readily study their own behaviors in this regard. Applied linguists, especially those situated more in the quantitative tradition, often borrow research ethics materials and ideas from other fields; however, it is not clear how relevant these materials are to applied linguistics. As a growing and evolving field, it can be argued that applied linguists should be having conversations about best practices within our own field.

As part of a three-year grant project<sup>1</sup> focused on issues related to quantitative humanities research, we are in the process of creating a set of research ethics training materials that could be used by researchers wanting to conduct quantitative research within the field of applied linguistics. These training materials should address valid instances of QRPs that scholars are likely to encounter. To achieve this goal, our training materials will be based on the results of a large-scale survey within the quantitative humanities that looks at how often researchers dealt with instances of QRPs and how severe they found each QRP to be.

As we began the process of creating the survey, we realized that there was not a field-specific list of QRPs created for the quantitative humanities nor quantitative (or general) linguistics. Instead of using a taxonomy of QRPs taken from another field or creating a taxonomy based on our own understanding of the topic, we selected a third option—to use a Delphi model of consensus (Dalkey & Helmer, 1963) to create a field specific taxonomy of QRPs. There are numerous training materials developed for the Delphi method (Day & Bobeva, 2005; Skulmoski et al., 2007; Turoff & Linstone, 2002) that sufficiently define the method and walk readers through the process of conducting such a study. Our goal in this paper is not to provide an in-depth, step-by-step tutorial of the Delphi method, but instead to showcase how the method works, to describe the decisions that we made along the way, and to hopefully provide readers with a basic understanding of how to conduct a Delphi study. In other words, this paper functions as a companion piece to other training documents, one that highlights complications and questions the approach. Reading this document along with the more robust training materials should prepare a scholar for conducting a Delphi study in applied linguistics with fewer bumps along the proverbial research path.

### ***What Is the Delphi Method?***

The Delphi method is an interactive and iterative data collection method that utilizes expert opinion to achieve a goal set out by the research team. It might be best to think of the Delphi method as a sequential focus group conducted via asynchronous communication methods.

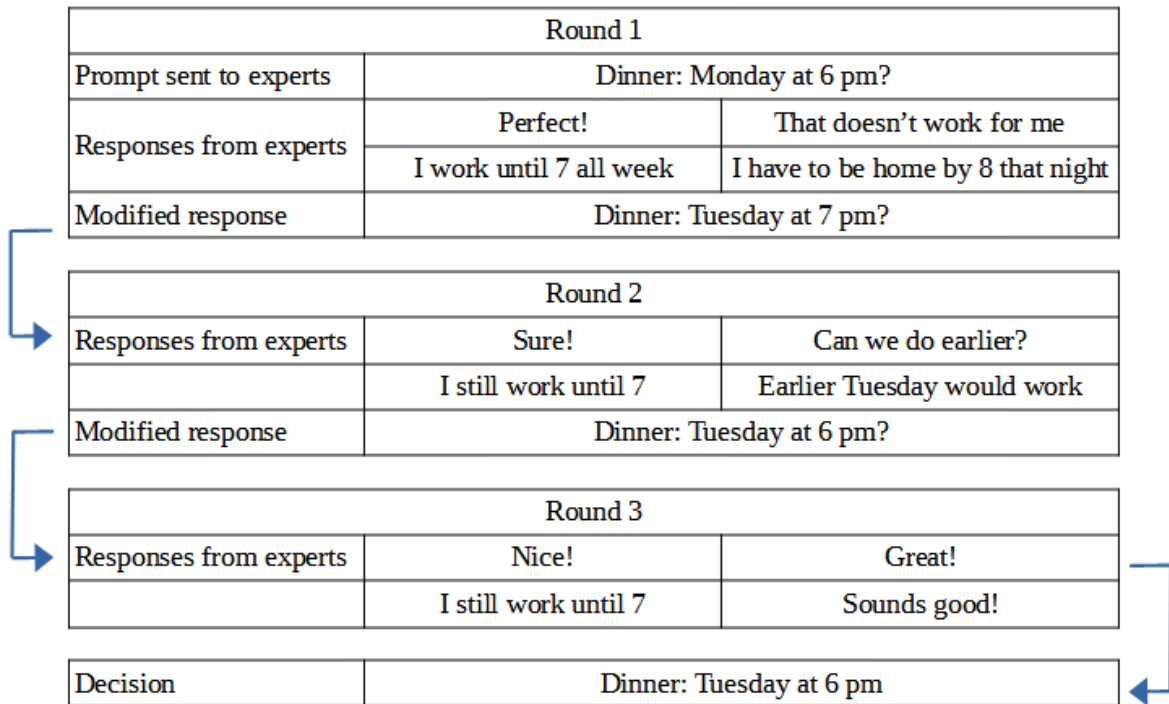
---

<sup>1</sup> *Questionable Research Practices: The (un)ethical handling of data in quantitative humanities research* (Larsson, Plonsky, Sterling, Kytö, Yaw, & Wood). Co-funded by the Swedish Research Council, the Bank of Sweden Tercentenary Foundation, and the Royal Swedish Academy of Letters, History and Antiquities (Project ID: FOE20-0017).

Typically, a panel of experts are sent a prompt and asked if they agree with it or not. If the panelists agree with the prompt, the prompt is said to have reached consensus and is set aside. If the panelists are unhappy with the prompt, they can make suggestions on how to improve it. All suggestions from all panelists are considered and the researcher attempts to modify the prompt to satisfy everyone. The modified prompt is sent back to the panelists who again provide feedback, and the researchers again attempt to re-modify the prompt if needed. This iterative process continues until the panelists are content with the prompt as written. The product of a Delphi study is a modified prompt that has reached sufficiently high agreement from a panel of experts.

A simple non-research example of the Delphi method in action is demonstrated in Figure 1. Here a group of friends (acting as experts) is trying to organize a group dinner. The person in the researcher role sends out a suggested time and day to meet. In round 1, one panelist is ok with the suggested day and time while the other three ask for modifications. At the end of round 1, the item has a 25% agreement rate and so modifications are required. The person acting as the researcher takes the feedback left by the panelists and attempts to create an optimal item. In our example, the researcher opts for a change in day and time to see if that improves the item. In round 2, there is one person who accepts the time while three ask for modified times. Based on the feedback, everyone appears to be happy with the day change and so a new time is proposed in the newly modified item. Finally, round 3 includes an earlier time and three people (75%) agree to the new time and date. The new prompt has reached consensus (as 75% was set as the threshold to pass) even though one person is excluded from the dinner. In this case, it is impossible to make everyone happy and so the item is revised in such a way as to make the greatest number of people accept the item. This simplified model represents the steps taken in a typical Delphi study, though often there are many prompts and more than four panelists. A more academic version of a Delphi method is included in Table 1.

**Figure 1**  
*Example of Iterative Delphi Process*



### ***A Brief History of the Delphi Method***

One of the first published examples using the Delphi method for obtaining expert consensus is from Dalkey and Helmer (1963), who shared findings from a classified experiment conducted in the 1950s using expert opinion to simulate Soviet selection of ideal industrial U.S. targets for bombing. Linstone and Turoff (2002) suggested that the Delphi method was particularly justified as a method in cases where 1) obtaining accurate information in other ways was impossible or prohibitively expensive, and/or 2) data modeling would be dominated by subjective inputs. Though the Delphi method's initial popularity was mainly confined to defense research and forecasting future events, this method also found early use as a tool for scientific developments (see Ament, 1970; Gordon & Helmer, 1964).

In relation to the present study, Delphi studies have also been used to identify types of scientific misconduct that would most strongly impact clinical trial results (Al-Marzouki et al., 2005). Indeed, the popularity of this research method may best be illustrated by the range of fields represented in primary studies using the Delphi method published within the last several years, including but not limited to: business (Ruiz-Pérez et al., 2021), cybersecurity (Chowdhury et al., 2022), dentistry (McLister et al., 2022), disaster management (Molassiotis et al., 2022), education (McGrath & Fischetti, 2021), fire safety (Thorp et al., 2022), international policy (Akartuna et al., 2022), medicine (Crofton et al., 2022; Yilmaz et al., 2022), nursing (Ye et al., 2022), optometry (Rodríguez-Zarzuelo et al., 2022), pharmacy (Olsen et al., 2021), resource management

(Breitenmoser et al., 2022), supply chain management (Peppel et al., 2022), and urban planning (Adu-McVie et al., 2021).

### ***The Delphi Method in Applied Linguistics Studies***

Despite the numerous recent examples of studies employing the Delphi method across a wide range of fields, its application in applied linguistics research has been more limited. A snapshot of this application is provided by Rodríguez-Lifante and Pereira (2021), who presented a Spanish-language review of the methodological and theoretical characteristics of 17 applied linguistics studies (written in either English or Spanish) that have used the Delphi method from 2014–2018. Their findings revealed that the majority of these studies were co-authored ( $n = 11$ , 65%) and published in English ( $n = 13$ , 76%). When selecting the expert panel, the number of experts ranged from 12 to 214 participants, with all studies reporting some attrition from the first to the final round (e.g., Uztosun, 2018). Some addressed this attrition by recruiting different experts for each round. The studies reviewed included one to four Delphi rounds, with two rounds being the most common ( $n = 8$ , 47%). The instruments used for data collection were primarily written and included Likert scales, open-ended questionnaires, interviews, and discussion groups. Methodologically, one weakness that Rodríguez-Lifante and Pereira noted among the publications reviewed was the lack of detailed reporting on the decision-making process involved, leaving readers with potential concerns about the validity and reliability of the instruments used, as well as the data collected. Moreover, a lack of transparency has implications for replication of Delphi research.

The Delphi method has been utilized in areas of applied linguistics that benefit from a structured approach to gathering expert input. On the pedagogical side, it has been used for investigating high-impact instructional practices for academic language development among English language learners (Pritchard & O'Hara, 2017), identifying core linguistic concepts for L1 grammar instruction (van Rijt & Coppen, 2017), and evaluating technology-mediated English language learning models (Abdullah & Siraj, 2018) as well as electronic learning materials (Hilinger & Rodríguez, 2019). In professional training, the Delphi method has been employed when developing models of professional competencies, such as for teachers using telecollaboration in the foreign language classroom (O'Dowd, 2015), instructors teaching in primary school settings in Turkey (Uztosun, 2018), and food science graduates communicating within their field (Weston et al., 2017). Additionally, the method has offered expert input on language assessment design and practices (Sevilla-Pavón et al., 2017). For translations, it has also been helpful (e.g., Amaral et al., 2016; Prazeres et al., 2015) at providing greater semantic accuracy, revealing which aspects of a translation have the most disagreement among translators, and highlighting cultural differences which impact how translators from diverse backgrounds interpret the original text (Rodríguez-Lifante & Pereira, 2021). The Delphi method has also offered insight on further issues of a cultural nature, such as how the media portrays intercultural communication (García-Jiménez et al., 2017) or how the values of religious messages relate to their transmission through the media (Rončáková, 2018). Overall, this method can be useful in building consensus on terminology, establishing

common criteria for evaluation, and validating empirical instruments (Rodríguez-Lifante & Pereira, 2021).

## **Applying the Delphi Method**

In this section of the paper, we detail the way that our research team conducted a Delphi study to illustrate application of the method. The Delphi method is flexible and shaped by the local context in which a study is conducted. We will highlight decisions we made and concerns we have/had in using this approach using our original study as a framing tool. In the end, we believe that the Delphi method allowed our grant project to start on strong footing and that the results of the Delphi method phase will make subsequent phases of our project more robust, generalizable, and valid for our intended audience.

### ***Research Team***

The true first step in any research project is to decide who will conduct the study. For a method such as Delphi, this step is critical since the research team will interpret and organize the responses of the expert panel. While the Delphi method is often presented as a community-driven approach, the reality is that this community voice is filtered through a panel of experts and then additionally filtered through a research team. The research team will organize ideas into themes and decode participants' comments. The research team not only has the final say in how to revise items, but also power in deciding what to research and how to ask questions that seed certain types of responses. Consequently, making sure that all team members are knowledgeable of the topic and the method are key to conducting a quality Delphi project.

For our particular project, the research team was put together prior to deciding to use the Delphi method. The team was deemed qualified to carry out this project because members had individually published papers, books, and other documents on research ethics, QRPs, research methodology, and have used the Delphi method on previous projects. We have also taught multiple research methods and statistics courses, served on ethical review panels, run conference strands on research methodology, and presented extensively on the topic at conferences. During our initial meetings, we brainstormed ideas for a project that met the grant call and settled on our three-phase approach (Delphi method, survey of QRPs, development of training materials). The team gathered resources on the method and read studies that had utilized the method from other fields. While the training materials and studies we compiled were normally of high quality, we felt that most lacked a true discussion of how a Delphi study functions in practice. As outlined above, one aim of this paper is to help the reader understand various strengths and weaknesses of the method, along with potential mistakes that we made during the process to support future scholars attempting to use this methodology.

### ***Research Aim***

After assembling a research team, the next stage in a Delphi study is to outline the aims and outcomes of the project. The Delphi method produces a community-generated artifact, but these artifacts can take on a host of different appearances from policy statements (Osborne et al., 2003) to making future predictions (Urias et al., 2020). Our aim was to catalog QRPs and build a taxonomy. We planned to use this taxonomy of QRPs to create a survey for a later study and to use our taxonomy and survey results in the creation of researcher training materials. As such, our outcome goal for the Delphi method portion was a list of QRPs worded in a way that we could easily transform into a survey. Thus, our artifact was a list of items in a loose structure. Our experts read and evaluated individual items but were not asked to structure the list itself.

### ***Experts***

The expert panel is the key to a quality Delphi study and much care should be devoted to ensuring that a panel consisting of experts with suitable expertise is selected. Of course, suitability will shift depending on the project as aims and usage of the Delphi method change. The Delphi method was a minimal portion of our overall grant project and was conducted to help provide stronger validity claims for our survey. Some of the decisions we made might not be ideal for studies that look to use the Delphi method to create artifacts or policies that have the ability to drastically influence the lives of people.

What makes a person an expert and how to select them shifts from project to project. An expert could be someone who is a member of a community or someone with specific content knowledge. In our project, we wanted scholars with experience in the quantitative humanities and with research ethics and/or QRPs. For our study, any scholar working in the quantitative humanities would technically qualify as an expert, as they would all have opinions on what exactly a QRP is and how such practices might make their way into the quantitative humanities. Since all quantitative humanities researchers had more or less equal status as experts, we needed to narrow down the list of eligible experts as best as possible. We attempted to find scholars who had published on research ethics and/or meta-research topics, or who displayed interest in the topic of QRPs. We rationalized that quantitative humanities scholars working in these particular areas would have more to say about the topic and would therefore be more useful informants for this project. In the end, we largely used a convenience sampling approach where each team member compiled their own separate list of scholars who they thought would be ideal expert candidates. Each team member's list differed in size and focus. During early meetings, we combined lists and discussed the known backgrounds of each candidate. Some names were retained because of their specialization (e.g., publishes on research ethics) while others were removed for individual reasons (e.g., was on sabbatical during data collection).

While our method of expert selection was deemed appropriate for our project, one of the reviewers for this manuscript suggested the use of a more systematic method, such as collecting and analyzing CVs for each scholar. This approach, or a modified version, would be helpful to use in certain Delphi studies and likely would have resulted in a different panel than the one we

constructed. However, this method would not have been conducive to our study. We would have had to collect documents on thousands of scholars (all scholars in all fields of the humanities). In the end, a decision must be made for how to compile a panel of experts and this should be done in a principled way so that any outcome will be viewed with confidence. How a team selects panelists would need to be included in the final report for a project, so that readers have the ability to evaluate the strength of the findings.

For a Delphi study, the number of experts is a variable that needs to be manageable. Including more people will allow for more diversity in voices and a sample that might be closer to representing the larger population. However, asking dozens of people to read and reread successive rounds of data collected and provide feedback can make the process challenging. On the other hand, having too few people runs the risk of missing opinions and resulting in an artifact that reaches consensus too quickly. It is important that each Delphi project finds its own ideal zone in terms of number and makeup of the expert panel. For us, the goldilocks zone was to have 10 panelists, as this number of people would provide us with a large enough sample to represent the overall community and their expertise would provide us with useful feedback.

The reader might have noticed a weakness of the Delphi method in the preceding paragraphs. Like many research methods, the Delphi method tends to lack true random sampling and has the potential to include bias in expert panel selection. For example, it is highly unlikely that someone conducting the exact same Delphi study would create an identical expert pool to ours. Another issue is the challenge of creating a representative sample of experts. The primary reason that a person is included as an expert is that they have expertise in a particular field and topic. Depending on the topic being investigated, the ability of the research team to consider factors such as race, gender, ethnicity, and others will vary. If there are only a handful of people in the world who fit the expertise requirement, it will be challenging to balance voices since those people would be the totality of the possible panelist pool. Having a completely balanced panel might not be possible but that does not mean that the research team should not try. It is always possible to search for people from different backgrounds or to limit the number of people who represent a portion of the population. For our project, we attempted to include a balance of gender and a representative number of Swedish/European vs. U.S.-based voices.

Another issue with the expert panel is that it consists of humans. Each iterative phase of a Delphi study compounds the number of chances for data to be omitted if a person cannot meet a deadline, or an expert might change their mind after something happens in their personal life. Since the Delphi method attempts to create an artifact from opinions, any changes to those opinions might impact the validity of the artifact. As an example, an expert on our panel might have had a negative IRB experience between rounds and thus viewed certain QRPs differently. This person might ask for a new change in round 5 that was not present in round 2. Additionally, our panel consisted of people from different cultures and countries who worked at different institutions with diverging schedules and backgrounds. Several rounds of data collection came near or during winter breaks. We also conducted the study during a substantial (as of the time of writing, the worst) spike in COVID-19 cases. Our panelists got sick, went on vacation, and were trying to wrap up their

semesters. It was not surprising to the research team that while nine people participated in round 1, only six participated in rounds 2.1 and 2.2. Eight participants took part in round 2.3, which occurred later in January, after most winter breaks were completed. Only one round of data collection included the full set of participants. This is quite common in Delphi studies, but it does mean that while we believe that our taxonomy is somewhat representative of the field of quantitative humanities, there were several rounds where not all voices were included.

## **Our Delphi Study**

In this section, we walk the reader through our Delphi study and include comments and issues that we faced along the way. While the details in this section are specific to our study, the reader should be able to understand the inner workings of the Delphi approach. To summarize the Delphi method, each round includes: sending a prompt to the experts, the experts reading and providing feedback to the prompt, and the research team making modifications to the prompt. The rounds are repeated using the modified prompt from the previous round until an item is accepted by the experts (because they like the prompt as it is written) or an item is discarded (because the experts either cannot agree or they agree to dislike the prompt).

### ***Data Collection: Round 1***

Like many research methods, the way data is collected in the Delphi approach has shifted over time due to changes in global communication. Older versions utilized pen and paper methods, often with experts sending responses through the mail. However, with advances in technology, the method can be completed using email, Internet-based survey programs, or other forms of asynchronous communication. We opted to use the Qualtrics online survey platform and emailed links to experts. Doing this allowed us to easily capture the data and export it directly into a useable form.

When conducting a Delphi study, the opening round of data collection is best thought of as the seed round (Linstone & Turoff, 2002). This is when experts provide more open information that the research team can use to create their target material. However, it is not required that the first round in a Delphi study act as a seed round. As an example, if a research project dealt with an already created policy statement that the researchers wanted to get an expert opinion on, it would make sense to send the document in question and get feedback on it. However, if the research team wanted to make a policy statement, it would be best to ask the panelists to contribute their ideas in a more open method.

For our study, we used six open-ended questions asking the panelists to contribute their thoughts on ethical decisions that researchers had to make on topics related to 1) funding, 2) data collection, 3) data analysis, 4) dissemination of results, 5) service to the field, and 6) mentorship. This phase was conducted using open textboxes in a Qualtrics survey. Panelists responded in lists and/or prose. One complication from our seed question round was that our questions might have

guided the panelists to respond in particular ways. We could have left the response extremely open by asking the experts to *please list all QRPs that you can think of*. However, it is not clear that such an open question would have yielded quality results either. Panelists might have only responded with QRPs considered more common, such as those related to statistics or publishing, and might have been less likely to include items related to funding or mentorship. While our seed round of questions potentially limited the responses from the panelists to only the areas we asked for, it is also possible that a restricted set of questions prompted them to consider different areas of QRPs.

The research team took the responses received from round 1 and attempted to create an organized taxonomy. We decided to drop questions 5 and 6 (service and mentorship) from the taxonomy of QRPs. The panel provided fewer responses to these questions and most of the responses either did not connect to research practices or were comments already made in response to other questions, such as issues related to publishing and funding. Finally, many of the comments included in these two questions connected more to the idea of being a good scholar, such as the need to provide all of one's own students with equal opportunities. While these comments were interesting and might be used in later material development, they did not seem to readily connect with the idea of questionable research so much as questionable scholarly activities.

We used the data in the remaining questions to formulate the first pass at a taxonomy of QRPs. The responses from the experts differed in style with some listing issues while others answered in full sentences. Since the responses were open in nature, experts also used different names for similar concepts. In order to create our initial set of taxonomy items, the first author took the responses provided by the experts and coded them using bottom-up, inductive methods. Themes emerged from this process, such as in the data analysis section where experts made statements such as "handling missing data and outliers" or "The researcher might play around with the data with different criteria for outliers, hoping that the results will be 'more favorable' one way or another." These comments were combined to form a theme labeled "methods of handling outliers" while the comment about missing data joined the theme of "lack of transparency in data cleaning." The rest of the research team then verified that each item mentioned by the experts was included in a theme. After creating a master list, the team rewrote each theme into survey style items. The theme of outliers was rewritten into the item "using unjustified methods of handling outliers" while the theme of cleaning data became the item "not being transparent with regard to the reporting on what steps were taken for data cleaning (e.g., removing cases/items without a stated criterion or justification for doing so)."

The themes themselves were at times not sufficient to create items and so the research team discussed our interpretations of the comments from the panelists. As an example, the theme of "selective selection of literature" resulted in taxonomy items such as "intentionally omitting relevant work because it does not align with one's theoretical or methodological approach" or "failing to refer to relevant work by other authors." Again, it is worth noting that the Delphi approach often leaves much room for the research team to insert their ideas into the process.

Finally, while the items generated by the experts and the research team covered considerable ground, we took one additional step to ensure that the taxonomy was not missing key items. We consulted existing taxonomies on QRPs from other fields (e.g., Kumar, 2008; Tauginienė et al., 2019) to see if there were any elements excluded by our panel of experts that seemed appropriate for our taxonomy. In total, 90% of our items came from the expert panel and the remaining items came from existing taxonomies and the research team. The low number of items added could indicate that our expert panel were able to cover most topics discussed in the QRP literature already.

Taken together, the three prongs of the initial creation of our QRP list—1) expert views, 2) research team suggestions, and 3) other taxonomies—provided us with broad coverage of QRPs. We are not under the assumption that we completely captured every QRP that might exist, but instead this tripartite approach increased our confidence that our list contained quality and content validity (i.e., adequate coverage of the target construct). As we continue with this project, we will undoubtedly add and subtract items.

### ***Rounds 2.1, 2.2, and 2.3***

In a Delphi study, the main set of rounds (for us rounds 2.1, 2.2, and 2.3) are used to refine the artifact being created. At the end of round 1, we had a taxonomy that included 62 items. We converted this list into a second Qualtrics survey. Each item was listed and contained three response options: 1) accept that an item was acceptable in its current form, 2) accept that the item was good enough but could use some minor revision, and 3) either reject the item completely or suggest major revisions to the item. The panelists were sent the survey link and asked to read each item. If the panelist judged the item to be acceptable as written, no additional action was required for that item. By selecting option 2 or 3, the experts acknowledged an item was not fully suitable for our final taxonomy and that revisions were required. Panelists were asked to provide written feedback for how we could modify the item so that it would be acceptable to them. This method was used in rounds 2.1, 2.2, and 2.3 with the only difference being that round 2.1 included a final textbox where the experts could add items not included in our initial list. No additional items were recommended.

The goal of rounds 2.1, 2.2 and 2.3 was to refine the taxonomy. In a Delphi study, it is not expected that every expert will agree with the others. Thus, a threshold score needs to be established at the start. If expert consensus surpasses this threshold, then an item would be deemed to have reached consensus. Our threshold was set at 80%, a somewhat arbitrary number, but having eight of ten academics agree seemed reasonable to the research team. One issue that surfaced during our project was that the threshold did not always match the number of respondents. We had a different number of experts respond in each round and never had the full 10-person panel that we had originally wanted. 80% of six people is 4.8 people and thus we were forced to decide to round up or down. We opted for the simple approach of rounding up if the 80% was 0.5 or higher and down if it was 0.4 or lower. Thus, in round 2.1 and 2.2 (each had six panelists respond) a total

of five experts rating an item as ‘acceptable’ was required. In round 2.3, eight panelists responded and so six experts were needed for an item to pass the threshold.

The experts were sent the Qualtrics survey that contained the working taxonomy and a set of directions for how to engage with the study. The experts rated each item and provided feedback if necessary. After we received all responses (or the deadline for responding passed) we analyzed their feedback. Once an item met our consensus rate of 80%, it was removed from our working list and included on the final taxonomy (see Plonksy et al, in preparation for the final taxonomy).

In round 2.1, 43 items met or surpassed the threshold score and thus were accepted with no need for modification. The remaining 18 items (the ones that did not receive 80% consensus) required modification. To modify the items, the research team looked at the experts’ comments and attempted to rewrite each item so that it satisfied the feedback provided. However, this was not always possible nor advisable. There were times when feedback clashed between experts and other times when the experts asked for changes that the research team believed inadvisable. An example of this process can be seen in Table 1. The research team spent several hours discussing each remaining item and its feedback. We suggested rewrites and at times decided that there was no way to fix an item. In cases when an item seemed impossible to change or when the feedback indicated that an item should be dropped, we did so. In the end, we modified the items and only 16 had yet to reach consensus in this round.

Figure 2 shows the path that our items took from round 1 to the completion of the Delphi method. By the end of round 1, our taxonomy included 62 items that were reviewed and rated by the experts. In round 2.1, 43 were accepted, 15 were modified, three were removed, and one additional item was included (an original item was broken into two new items). In round 2.2, nine items were accepted, six were modified, and one was removed. In round 2.2, nine items were accepted, six were modified, and one was removed. In the final round, all six items were accepted, thus eliminating the need for any further rounds.

**Figure 2**  
*Path of Items through the Delphi Model*

	R2.1	R2.2	R2.3	Final
Accepted	43	9	6	58
Revised	15	6		
Removed	<del>3</del>	<del>1</del>		<del>4</del>
Added	1	0		
Total	62	16	6	

**Table 1***Example of Item Revisions by Round*

Round	Item	Example comment	Round decision
2.1	Nudging participants to join a study using monetary (or other) incentives	“Need to better define nudging. Is it a QRP if we all do it?”	Revise
2.2	Recruiting participants to join a study in a way that saying no feels wrong (e.g., using an emotional story during classroom recruitment)	“Maybe change to 'in a way that makes refusal difficult or uncomfortable' 'Emotional story'....hmmm, I'm all in favor of emotion. How about '..using coercion during classroom recruitment, or by pleading or shaming participants into agreeing)'?”	Revise
2.3	Recruiting participants to join a study in a way that makes refusal difficult or uncomfortable	“...I might add some kind of intensifier/qualifier before "difficult or uncomfortable" when thinking about this more - "unduly" might work. ...”	Accept

Table 1 shows how one item was revised based on expert feedback. While this example only includes one instance of feedback per round, most items included the opinions of several experts that needed to be balanced. Our overall idea for this example item was to depict a situation in which a researcher recruiting participants for their study might go into a classroom and not only read their promotional materials but also tug at the heartstrings of potential participants. The researcher could potentially say something like “I really need you all to participate in this research because if you don’t, I won’t be able to graduate on time and might lose my funding.” This form of nudging is a grey area as there was no ‘force’ or ‘coercion’ applied, but it is instead similar to emotional blackmailing. In round 2.1, we included the research jargon term ‘nudge’ which was potentially unknown by at least one panel member. This person’s comments asked for a better definition of the term. We decided to modify this item and used a synonym for the term instead. In round 2.2, the feedback suggested that we use ‘coerce’ which was a term we wanted to avoid since coercion has a special definition in research ethics and using methods that coerce people would normally be considered unethical, not simply questionable. We thus opted not to follow this particular expert's recommendation on the topic. However, we did use their exact phrasing of “in a way that makes refusal difficult or uncomfortable” for round 2.3. In this final round, the item

was finally rated ‘accept’ by enough panelists to pass our threshold. That said, there was a final comment asking us to use the term ‘unduly.’ Since the item had passed threshold, we did not feel the need to modify it further.

As can be seen in our breakdown of the Delphi method, the method works in iterative steps. Each round should have fewer and fewer items and each item should be closer to the ideal wording the expert panel wants. At the outset of a Delphi project, it is not possible to know how many rounds of revision will be required. There is a chance, though small, that all of the experts will agree with the initial items and only one round will be required. However, depending on the number of prompts, the nature of the topic, the experts themselves, and more, there could be an infinite number of rounds possible. Our project reached complete consensus in three rounds of revision, which seems standard in comparison to other projects using the Delphi method (Rodríguez-Lifante & Pereira, 2021).

Going through four rounds allowed us to create a list of QRPs that was seeded by the experts and modified by the research team. The three rating rounds allowed us to refine items so that they matched the expectations of the experts. In the end, we were left with a list of QRPs that were mostly generated by experts, rewritten by skilled researchers, matched previous literature, and then improved upon by a panel of experts representing a broader research community.

## **Discussion**

After completing the Delphi process, the team met to discuss the extent to which we were satisfied with the method and whether we would use it for future projects. We will share some of our insights here in hopes that it will be helpful for researchers wishing to apply the method to their projects. While not the focus of this paper, we highly suggest reflective discussion for all research teams considering this method.

A question asked by a reviewer of this manuscript was if our Delphi project produced a taxonomy of QRPs that differed from existing versions. Our answer is not straightforward. Applied linguists (at least the experts included in our project) largely included similar items found in other taxonomies of QRPs. However, it was not known at the start of our Delphi project that this would be the case. Even though the content of our taxonomy is similar to other taxonomies cited in this paper, the items themselves were written and modified in ways that made sense to our field in particular. Through three rounds of revision, we were able to produce items that should be simpler for applied linguists to read and understand. Had we used items copied from another taxonomy or created the items ourselves, results might have been skewed simply based on item wording. In the end, while the content of our taxonomy is similar to others, from using an evidence-based approach we now have greater confidence that each item in our taxonomy and survey not only belong, but also are written in a way that is appropriate for quantitative humanities research. However, like most products, we will not know the extent of the readability of our items until we complete the

next phase of the project where we send the survey (created from the taxonomy items) to the field at large.

One clear downside of this method is that it is time consuming. Had we created a survey from an existing taxonomy of QRPs, we likely would have been able to publish the results with little difficulty by standing on the well-trodden grounds of claiming that our materials were valid because someone else had created and validated the taxonomy used. Using an existing taxonomy would have been easy and efficient. In fact, we probably could have conducted the entire survey project and submitted a write-up for publication in the time that we focused only on the Delphi portion. The Delphi process took months to complete and used time, energy, and grant resources.

The main reason that we found the Delphi method useful was that it allowed us to develop a community-driven taxonomy of items, which was our original goal. The process of using a community—in this case represented by the panel of experts—to generate research materials bolsters the results with a sense of both content validity and usefulness to a particular audience. Often, validity is assumed if an item was used in a previous study. The fact that materials are internally consistent is generally preferable, yet consistency is not always critical. A survey might result in a certain number of factor loadings, which is important, but internal validity does not indicate that the questions asked were valuable.

The iterative process of the Delphi method provided us opportunities to make corrections as we developed our list of QRPs. Most items in round 2.1 were based on ideas from the expert panel but filtered through the research team. It is certainly possible that our current taxonomy is missing items. Science itself is an iterative process; however, we hope that our materials go through additional refinement in order to best match the ‘reality’ of QRPs. The three-pronged approach of using experts, literature, and the research team provides the field with a useful tool moving forward.

As with all research methods, there are several issues to using the Delphi method and many of those issues were pointed out throughout this paper. No method is going to be perfect for all situations and all methods will have flaws. In our version of the Delphi method, we constricted the type of QRPs by using six initial questions. We could have instead used one open question such as “write out all the QRPs that you can think of”. Other issues, such as the arbitrary nature of asking for 80% agreement, are unavoidable. Using any metric of agreement (be it 1% or 100%) is arbitrary. Anyone who has attempted to get a group of academics to pick a restaurant at a conference might be impressed that we were able to get 80% of our experts to agree on anything!

### ***Delphi Method Versus Focus Group***

A topic that we discussed during our summary meeting was whether it would have been simpler to host a focus group. The Delphi method can be thought of as an asynchronous focus group, and so comparing it to a focus group seems apt. However, there are notable differences, and we believe that the Delphi method has some advantages over focus groups, which would justify having both methods remain in our researcher toolkits.

One consideration is the time factor. While both methods require copious amounts of time to organize, plan, and set up, there are key differences in the active time each takes. Best practices for focus groups suggest keeping the time limit to around one hour (Morgan & Krueger, 1997). Each round of the Delphi study took one to two weeks to complete. Best practices for both methods suggest keeping the number of participants low, though Delphi studies tend to work with more individuals than focus groups. Finding time to conduct an international focus group might have been complex. Trying to engage and coordinate with 10 academics, all of whom are advanced enough in their careers to be considered experts on a topic, plus including at least one member of the research team in a single video call might have been a Lovecraftian-fueled nightmare in organization.

One comparison between focus groups and Delphi studies that might not be obvious at first is the required methodological expertise of the researcher. Focus groups are hard to conduct and require practice to become proficient in. The interviewer needs to monitor questions, potentially ask follow-up questions, ensure that all participants are provided a chance to engage with the questions, steer the discussion toward answering the question that was actually asked, and stay out of the process as much as possible. A Delphi study on the other hand is not very complex to run. After the materials have been established, the researcher sends out correspondence, creates a (survey) platform for experts to use, and potentially answers basic questions. The analysis phase of a Delphi study is also simple to conduct. The material is organized as a question/answer response pair and is likely already in text form. Focus group data is messy and can be very time consuming to transcribe, especially if members of the group talk over each other. Depending on the needs of the research project, it might make sense for a researcher to use a method with a lower floor of entry such as the Delphi method. On the other hand, depending on the depth of knowledge required, one might want to utilize a focus group.

The Delphi method also provides a level of anonymity that focus groups do not. Our experts were able to expatiate on a number of risky topics (depending on the QRP admitting to one might have serious career complications, whether real or simply perceived). The Delphi method allowed the experts, who were unaware of who else was serving on the panel, to potentially be more truthful in their responses.

Another key difference between the two methods is the type of interaction. In a focus group, it is not possible to ensure that everyone is able to respond nor given equal talk time. The need to complete a series of questions in a limited timeframe, the fact that some people might talk for extended periods of time, and other factors can severely limit an individual participant's contribution. The Delphi method allows each person to respond at their own pace without the constraints placed on normal group interaction. However, while we might look at the Delphi method's isolation of experts as a good thing in the context of our project, it did mean that 'group-think' was not possible. In a focus group, participants can generate ideas that might spur another person to think differently. Had each person read the comments from the whole panel during each round, they might have thought differently of the topic or had different types of input to provide.

### *Implications*

Science is not about getting the right answer the first time, but instead is about refining our knowledge until we feel confident that we know enough to move on to other questions. The Delphi method itself is a good model of the scientific thought process due to its iterative nature. The following are three project types in which we could see a Delphi approach being useful to the field of applied linguistics. These and similar projects would benefit from the iterative process and the community-driven nature of knowledge creation involved within the Delphi method.

- **Group consensus:** Applied linguistics often lacks a consensus on very basic topics (what is language, what is proficiency, what is applied linguistics). Arguments could be built from reading the literature but even then, it is hard to tell how much agreement the field has. A Delphi study could attempt to define some of the terms that we often use but do not seem to agree on, providing the field with a baseline understanding of our terminology.
- **Policy documents:** Delphi studies have often been used for the creation of policy documents and this idea could be brought into applied linguistics. Documents related to early language programs, language revitalization projects, or best-practices for the use of language on museum placards might benefit from this approach.
- **Project creation:** Often the research process begins with a researcher having an idea or question about a particular topic and then setting out to investigate the answer. Instead, the Delphi approach might be useful in bringing in stakeholders to the research process early and allowing them to help guide the research question and topic. Doing so might allow applied linguists to better support the needs of stakeholders and to make our research more visible and useful.

In the end, would we recommend other scholars use the Delphi method in their research? Given the needs of the research project, we would answer with a resounding ‘yes’. Any method that provides applied linguistics materials with additional validity arguments is worth considering. The Delphi method also has uses that align well with other meta-research initiatives within the field. In this paper, we have attempted to outline how we conducted a Delphi project but at the same time highlight the method’s various benefits and drawbacks.

## References

- Abdullah, M. R. T. L., & Siraj, S. (2018). The evaluation of mLearning implementation of model for English language learning via fuzzy Delphi method. *Pertanika Journal-Social Sciences y Humanities, Selangor*, 26, 33-54. <http://psasir.upm.edu.my/id/eprint/13734/1/JSSH%20Vol.%2026%20%28S%29%20Jan.%202018%20%28View%20Full%20Journal%29.pdf#page=47>
- Adu-McVie, R., Yigitcanlar, T., Erol, I., & Xia, B. (2021). Classifying innovation districts: Delphi validation of a multidimensional framework. *Land Use Policy*, 111, 105779. <https://doi.org/10.1016/j.landusepol.2021.105779>
- Akartuna, E. A., Johnson, S. D., & Thornton, A. (2022). Preventing the money laundering and terrorist financing risks of emerging technologies: An international policy Delphi study. *Technological Forecasting and Social Change*, 179, 121632. <https://doi.org/10.1016/j.techfore.2022.121632>
- Al-Marzouki, S., Roberts, I., Marshall, T., & Evans, S. (2005). The effect of scientific misconduct on the results of clinical trials: A Delphi survey. *Contemporary Clinical Trials*, 26, 331-337. <https://doi.org/10.1016/j.cct.2005.01.011>
- Amaral, A. B. C., Rider, E. A., Lajolo, P. P., Tone, L. G., Pinto, R. M., Lajolo, M. P., & Calhoun, A. W. (2016). Development of a Brazilian Portuguese adapted version of the Gap-Kalamazoo communication skills assessment form. *International Journal of Medical Education*, 7, 400-405. <https://doi.org/10.5116/ijme.583a.df42>
- Ament, R. H. (1970). Comparison of Delphi forecasting studies in 1964 and 1969. *Futures*, 2(1), 35-44. [https://doi.org/10.1016/0016-3287\(70\)90006-6](https://doi.org/10.1016/0016-3287(70)90006-6)
- Arroyo González, M. J., & Berzosa Ramos, I. (2018). Educational attention to immigrant students: In search of consensus. *Revista de Educación*, 379, 181-203. <https://doi.org/10.4438/1988-592X-RE-2017-379-367>
- Assenova, R. S., Le Reste, J. Y., Foreva, G. H., Mileva, D. S., Czachowski, S., Sowinska, A., Nabbe, P., Argyriadou, S., Lazic, D., Hasaganic, M., Lingner, H., Lygidakis, C., Munoz, M.-A., Claveria, A., Doerr, C., Van Marwijk, H., Van Royen, P., & Lietard, C. (2015). Cross-cultural validation of the definition of multimorbidity in the Bulgarian language. *Folia Medica*, 57(2), 127-132. <https://doi.org/10.1515/folmed-215-0031>
- Breitenmoser, L., Quesada, G. C., Anshuman, N., Bassi, N., Dkhar, N. B., Phukan, M., Kumar, S., Babu, A. N., Kierstein, A., Campling, P., & Hooijmans, C. M. (2022). Perceived drivers and barriers in the governance of wastewater treatment and reuse in India: Insights from a two-round Delphi study. *Resources, Conservation and Recycling*, 182, 106285. <https://doi.org/10.1016/j.resconrec.2022.106285>
- Chowdhury, N., Katsikas, S., & Gkioulos, V. (2022). Modeling effective cybersecurity training frameworks: A Delphi method-based study. *Computers & Security*, 113, 102551. <https://doi.org/10.1016/j.cose.2021.102551>

- Committee on Science Engineering and Public Policy (U.S.). Panel on Scientific Responsibility and the Conduct of Research. (1992). *Responsible science: Ensuring the integrity of the research process*. Washington: National Academy Press.
- Crofton, E., Meredith, P., Gray, P., & Strong, J. (2022). Use of sensory modulation approaches to improve compression garment adherence in adults post-burn: An e-Delphi study. *Burns*. Advance online publication. <https://doi.org/10.1016/j.burns.2022.04.001>
- Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9(3), 458-467. <https://doi.org/10.1287/mnsc.9.3.458>
- Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of Delphi studies. *Electronic Journal of Business Research Methods*, 3(2), 103-116.
- García-Jiménez, L., Rodrigo-Alsina, M., & Pineda, A. (2017). The social construction of intercultural communication: A Delphi study. *Journal of Intercultural Studies*, 38(2), 228-244. <https://doi.org/10.1080/07256868.2017.1289903>
- Gordon, T.J., & Helmer, O. (1964). *Report on a long-range forecasting study* (Paper P 2982). The RAND Corporation. <https://apps.dtic.mil/sti/pdfs/AD0607777.pdf>
- Hilinger, A., & Rodríguez, V. P. (2019). Análisis comparativo y evaluación de la calidad de los recursos electrónicos para aprender lenguas extranjeras: el caso de Cambridge English y Aveteca. *Edmetic*, 8(1), 36-54. <https://doi.org/10.21071/edmetic.v8i1.11156>
- Hsin, M. C., Chien, S. P., Hsu, Y. S., Lin, C. Y., & Yore, L. D. (2016). Development and validation of a Taiwanese communication progression in science education. *International Journal of Science and Mathematics Education*, 14(1), 125-143. <https://doi.org/10.1007/s10763-014-9589-y>
- Isbell, D. R., Brown, D., Chen, M., Derrick, D. J., Ghanem, R., Arvizu, M. N. G., Schnur, E., Zhang, M., & Plonsky, L. (2022). Misconduct and Questionable Research Practices: The Ethics of Quantitative Data Handling and Reporting in Applied Linguistics. *The Modern Language Journal*, 106(1), 172–195. <https://doi.org/10.1111/modl.12760>
- Kumar, M. N. (2008). A review of the types of scientific misconduct in biomedical research. *Journal of Academic Ethics*, 6, 211-228. <https://doi.org/10.1007/s10805-008-9068-6>
- Linstone, H. A., & Turoff, M. (Eds.). (1975, 2002). *The Delphi method*. Addison-Wesley. <https://web.njit.edu/~turoff/pubs/Delphibook/index.html>
- McGrath, J., & Fischetti, J. (2021). The future of compulsory schooling: Participant developed scenarios from a modified Delphi survey. *Futures*, 133, 102818. <https://doi.org/10.1016/j.futures.2021.102818>
- McLister, C., Moore, C., Harkness, S. M., O'Neill, C., Donnelly, M., & McKenna, G. (2022). Appropriateness of tooth replacement strategies for adult patients in the United Kingdom with reduced dentitions—a modified Delphi analysis. *Journal of Dentistry*, 104125. <https://doi.org/10.1016/j.jdent.2022.104125>
- Mirbagheri, S. A. (2014). A knowledge management view of teaching English as a foreign language (TEFL) in General Educational System (GES) of Iran. *Procedia-Social and Behavioral Sciences*, 136, 148-152. <https://doi.org/10.1016/j.sbspro.2014.05.305>

- Molassiotis, A., Guo, C., West, C., & Loke, A. Y. (2022). Identifying the priorities in nursing research on disaster management: A Delphi study and international survey. *International Journal of Disaster Risk Reduction*, 102935. <https://doi.org/10.1016/j.ijdr.2022.102935>
- Morgan, D., & Krueger, R. (Eds.). (1997). *The Focus Group Kit*. Sage Publishing
- Mustapha, N. H., Rouyan, N. M., Mustapha, N. F., Daud, N., & Mat, K. S. (2018). Expert Views on Metacognitive Strategies for Arabic Language Learning: An Application of the Fuzzy Delphi Method. *International Journal of Engineering & Technology*, 7(4.38), 1034-1038.
- O'Dowd, R. (2015). Supporting in-service language educators in learning to telecollaborate. *Language Learning & Technology*, 19(1), 63-82. <https://llt.msu.edu/issues/february2015/odowd.pdf>
- Olsen, A. A., Wolcott, M. D., Haines, S. T., Janke, K. K., & McLaughlin, J. E. (2021). How to use the Delphi method to aid in decision making and build consensus in pharmacy education. *Currents in Pharmacy Teaching and Learning*, 13(10), 1376-1385. <https://doi.org/10.1016/j.cptl.2021.07.018>
- Osborne, J., Collins, S., Ratcliffe, M., Millar, R., & Duschl, R. (2003). What “ideas-about-science” should be taught in school science? A Delphi study of the expert community. *Journal of Research in Science Teaching*, 40(7), 692–720. <https://doi.org/10.1002/tea.10105>
- Peppel, M., Ringbeck, J., & Spinler, S. (2022). How will last-mile delivery be shaped in 2040? A Delphi-based scenario study. *Technological Forecasting and Social Change*, 177, 121493. <https://doi.org/10.1016/j.techfore.2022.121493>
- Plonsky, L., Larsson, T., Sterling, S., Kytö, M., & Yaw, K. (in preparation). Developing a taxonomy of ethical decisions in applied linguistics research. In De Costa, P. I., Rabie-Ahmed, A., & Cinaglia, C. (Eds.), *Ethical issues in applied linguistics scholarship*. John Benjamins
- Prazeres, F., Santiago, L. M., & Simões, J. A. (2015). Defining multimorbidity: from English to Portuguese using a Delphi technique. *BioMed Research International*, 2015, 965025. <https://doi.org/10.1155/2015/965025>
- Pritchard, R., & O'Hara, S. (2017). Framing the teaching of academic language to English learners: A Delphi study of expert consensus. *TESOL Quarterly*, 51(2), 418-428. <https://doi.org/10.1002/tesq.337>
- Rodríguez-Lifante, A., & Pereira, M. M. B. (2021). The Delphi method in applied linguistics in the light of a theoretical and critical analysis. *Revista Brasileira de Linguística Aplicada*, 21(1), 271-293. <https://doi.org/10.1590/1984-6398202116351>
- Rodríguez-Zarzuelo, G., Gómez-Niño, Á., & Martín-Herranz, R. (2022). A Delphi study to identify and assess professional competencies in the education of optometrists. *Journal of Optometry*. Advance online publication. <https://doi.org/10.1016/j.optom.2021.12.005>
- Rončáková, T. (2018). Religious messages and the media code – Inherent features of the media language code and the transmission of religious messages. *KOME: An International Journal of Pure Communication Inquiry*, 5(2), 19-37. <https://doi.org/10.17646/KOME.2017.22>

- Ruiz-Pérez, F., Lleo, A., & Ormazabal, M. (2021). Employee sustainable behaviors and their relationship with Corporate Sustainability: A Delphi study. *Journal of Cleaner Production*, 329, 129742. <https://doi.org/10.1016/j.jclepro.2021.129742>
- Sevilla-Pavón, A., Gimeno-Sanz, A., & García-Laborda, J. (2017). Actitudes docentes hacia los ejercicios de la Prueba de Acceso a la Universidad informatizada. *Educação e Pesquisa*, 43(4), 1179-1200. <https://doi.org/10.1590/S1517-9702201612149283>
- Skulmoski, G., T. Hartman, F., & Krahn, J. (2007). The Delphi Method for Graduate Research. *Journal of Information Technology Education: Research*, 6, 001–021. <https://doi.org/10.28945/199>
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and engineering ethics*, 12(1), 53–74
- Sterling, S., & Gass, S. (2017). Exploring the boundaries of research ethics: Perceptions of ethics and ethical behaviors in applied linguistics research. *System*, 70, 50–62. <https://doi.org/10.1016/j.system.2017.08.010>
- Tauginienė, L., Gaižauskaitė, I., Razi, S., Glendinning, I., Sivasubramaniam, S., Marino, F., Cosentino, M., Anohina-Naumeca, A., & Kravjar, J. (2019). Enhancing the taxonomies relating to academic integrity and misconduct. *Journal of Academic Ethics*, 17(4), 345-361. <https://doi.org/10.1007/s10805-019-09342-4>
- Thorp, T. N., Fischer, E. C., & Sinha, A. (2022). A Delphi study to identify and prioritize research gaps for the incorporation of a fire into life cycle assessment of structures. *Fire Safety Journal*, 129, 103571. <https://doi.org/10.1016/j.firesaf.2022.103571>
- Turoff, M., & Linstone, H. A. (2002). The Delphi method-techniques and applications. <https://web.njit.edu/~turoff/pubs/DelphiBook/index.html>
- Urias, E., Vogels, F., Yalcin, S., Malagrida, R., Steinhaus, N., & Zweekhorst, M. (2020). A framework for Science Shop processes: Results of a modified Delphi study. *Futures*, 123, 102613. <https://doi.org/10.1016/j.futures.2020.102613>
- Uztosun, M. S. (2018). Professional competences to teach English at primary schools in Turkey: A Delphi study. *European Journal of Teacher Education*, 41(4), 549-565. <https://doi.org/10.1080/02619768.2018.1472569>
- Van Rijt, J., & Coppen, P. A. (2017). Bridging the gap between linguistic theory and L1 grammar education—experts’ views on essential linguistic concepts. *Language Awareness*, 26(4), 360-380. <https://doi.org/10.1080/09658416.2017.1410552>
- Weston, E., Crilly, J., Mossop, L., & Foster, T. (2017). Competencies for food graduate careers: developing a language tool. *Higher Education Pedagogies*, 2(1), 101-115. <https://doi.org/10.1080/23752696.2017.1366275>
- Ye, J., Tao, W., Yang, L., Xu, Y., Zhou, N., & Wang, J. (2022). Developing core competencies for clinical nurse educators: An e-Delphi-study. *Nurse Education Today*, 109, 105217. <https://doi.org/10.1016/j.nedt.2021.105217>
- Yaw, K., Plonsky, L., Larsson, T., Sterling, S., & Kytö, M. (under review). Research Ethics in Applied Linguistics.

Yilmaz, Y., Chan, T. M., Thoma, B., Luc, J. G., Haas, M., Preiksaitis, C., Tran, V., & Gottlieb, M. (2022). Identifying social media competencies for health professionals: An international modified Delphi study to determine consensus for curricular design. *Annals of Emergency Medicine*. Advance online publication. <https://doi.org/10.1016/j.annemergmed.2022.02.016>