Exploring Standard-Based Policy Updates Using a Modified Policy Delphi: "The Categorical Delphi Technique"

Valentine Olusegun Matthews[1]

[1] Lancaster University, UK

Correspondence: Valentine Olusegun Matthews, Lancaster University, UK.

Received: August 6, 2025 Accepted: October 22, 2025 Online Published: November 3, 2025

doi:10.5430/ijhe.v14n6p1 URL: https://doi.org/10.5430/ijhe.v14n6p1

Abstract

The Delphi technique is a systematic method for evaluating anonymized expert opinions to achieve convergence on complex issues. As a modified Classical Delphi, Policy Delphi explores diverse relationships between policy options rather than forcing consensus. This study explores the idea of "Categorical Delphi", a novel adaptation of Policy Delphi.

The research evaluates categorical relationships between regulatory policy and its addendum policy metrics. It also explores practitioners' attribution of quality and value of the metrics of the addendum policy within the regulatory environment. Predefined categorical relationships between two policies are established to guide the experts' reflection on the updates made by an addendum policy on a primary policy's metrics. Categorical Delphi employs multiple rounds of questionnaires increasingly refining their positional consensus. Attributions of quality of the metrics and the relative importance are determined by rating and ranking the quality of the addendum metrics.

The Categorical Delphi Technique effectively evaluated the categorical relationships between primary and addendum policies. The addendum metrics were perceived by faculty and managers as moderate-to-high quality additions to the quality environment. The Categorical Delphi Technique offers a robust approach for evaluating categorical relationships between a regulatory policy and a policy addendum, providing actionable insights for institutions.

Keywords: Categorical Delphi, Policy Delphi, policy evaluation, regulatory policies, higher education, expert consensus

1. Introduction

1.1 Introduction

Policies are authoritative allocations of values (Easton, 1965). Higher education Quality Assurance (QA) regulatory policies are allocations of the values of expectations to which institutions comply. Policies like the Bologna process have expedited the introduction and elaboration of institutionalized QA and quality management mechanisms' (Seyfried and Pohlenz, 2018). Quality within the higher education sector could be viewed as a 'conceptual tool through which [QA is] implemented'. Although quality controls and QA are aspects of quality management systems, QA is implemented to 'verify that quality controls are being maintained' (Padró, 2009, p.98). National regulatory frameworks are foundational to the development of internal QA in universities (IIIEP, 2018). Although external and internal QA are widely accepted, it is viewed as a bureaucratic burden and an illegitimate managerial approach to quality (Seyfried and Pohlenz, 2018). Nonetheless, internal and external QA ensure that Higher Education Institutions (HEI's) comply with national regulatory standards.

The massification and internationalization of tertiary education has increased awareness of the relevance of QA (Bollaert, 2014). Thus, the achievement of outcomes and accountability are integral to university managers' daily work (Broucker and De Witt, 2015; Van Vught and De Boer, 2015). Internal quality managers interpret regulatory requirements to ensure institutional compliance of HEIs with the primary intent of regulatory requirements. HEIs seek to decipher the metrics of the regulatory addenda. This study aims to develop an approach that can be used to explore practitioners' collective experiences within HEIs to evaluate the updates made by a regulatory policy addenda's metrics on the primary policy.

1.2 What is the Research Problem?

It is not unusual for national regulators to provide policy addenda. For example, the UAE's Compliance Inspection Framework" (CIF) (2020) is a regulatory policy addendum designed to update the provisions of the primary higher regulatory framework - Commission for Academic Accreditation's Standard (CAA) (2019). Seamless implementation of both policies without repetition is the challenge for the HEIs.

There is no 'recipe for carrying out policy analysis in education' (Ozga, 2000). The appropriate approach employed in education policy analysis depends on the nature of the policy being analyzed' … and 'the purposes of policy analysis is equally important for the theoretical and methodological approaches to be adopted' (Rizvi & Lingard, 2010). Policy analysis in this research follows the traditional definition of an "Analysis of Policy" and not an "Analysis for Policy" because the policy is already created by national regulators. Cibulka (1994) saw both categories as sitting on several levels of the applied education policy studies continuum. This study explores how a further modification of Policy Delphi analysis the categorical update types made by the addendum policy metrics on the primary policy. The research questions were as follows.

RQ1. How can the Categorical Delphi Technique be used to explore improvements made by an addendum policy to higher education regulatory framework?

RQ2. Does the Categorical Delphi Technique decipher the "update types" made by addendum policy metrics on the regulatory standard?

RQ3. How do quality managers rate the "quality attributes" of the addendum metrics?

2. Literature Review

The Delphi technique makes a difference in synthesizing the perspectives of a collective of experts; thus, it is useful for interpreting ambiguous or evolving policy updates. 'Problems linked to group communication and decisions that lend themselves to the use of group involvement' are seen as appropriate to be explored by the Delphi technique (Linstone and Turoff, 1975, p.11). The Delphi technique has been used to explore or expose underlying assumptions, cohere conceptions, or a topic and to understand the diverse and interrelated aspects of a research question (Hasson et al., 2000). This suggests that by using the Delhi technique, the divergence between the two policies could be identified.

Delphi techniques excel in the employment of anonymous iterative rounds of questionnaires. This approach minimizes bias from dominant participants or group thinking, fostering independent and reflective input (Linstone & Turoff, 1975). The flexibility of the method allows for the customization of specific policy domains, such as education policy. A weakness of this technique is the introduction of bias from reliance on experts. The ability of the panel to critically express their diverse opinions or perspectives; their diverse expertise, ideology, or non-inclusion of representation of stakeholders; and the resulting interpretation of policy updates may be skewed (Hsu & Sandford, 2007).

Modifications are sometimes made to the Classical Delphi technique by combining it with methodologies such as 'focus groups, interviews, or results of a review' (Keeney, 2010, p.229). The Delphi technique is typically adopted because it allows consensus of opinions on concerns (Carrol, 2004). On the other hand, it has been suggested that consensus is necessary for the success of policies, policy implementation, policy reforms and policy translation into effective change (Santiago et al., 2008). Achievement of consensus on a complex issue is characterized by uncertainty or a lack of empirical evidence (Powell, 2003). Consensus is usually the focus of Delphi approaches (Hasson et al., 2000; Turoff and Linstone, 2002) and the Policy Delphi technique" focuses on the exploration of policy concerns (Turoff, 1970). The true nature of the policy concerns should be evaluated and known by the practitioners for its effective implementation.

2.1 Justifications for the Categorical Delphi Technique?

Studies on the impact of education policies and practices often focus on methodological issues, such as the identification of econometric methods for measuring policy impacts (Schlotter et. al., 2009). Categorical Delphi uses expert knowledge to evaluate the relationship between metrics of two policies using categorically defined update types. This allows for nuanced comparisons that expose dissent within categorical relationships. The Policy Delphi is not designed to force consensus (Hsu & Sandford, 2007) however, the Categorical Delphi seeks consensus. Hence, the quality of the outcomes of the Categorical Delphi is tied to how effectively categorical relationships are predetermined. It is valuable for analyzing complex interactions between two or more higher education regulatory standards. Like policy analysis, it builds on participants' anonymity, reducing bias from dominant experts, enabling

rigid categorization of policy differences, and reducing power dynamics in regulatory debates. Categorical Delphi analyses the variances between the two policies by requiring knowledgeable experts from both policy documents to evaluate the variations between the policy metrics into categorical relationship types. Thus, it can be used to identify how an addendum policy updates a pre-existing policy.

Analysis of the outcomes of the Categorical Delphi is quicker than the Classical Delphi or Policy Delphi techniques. Categorical Delphi transfers responsibility for the evaluation of the primary and addendum policies and for evaluating the relationship between the primary and addendum policies to experts. Thus, it evaluates policy relationships using expert reflections of categorical policy links. Within the context of this study, the categorical relationships include the addendum metrics are (i) wholly pre-existing in the primary policy, (ii) partially existing in the primary policy, or (iii) entirely new additions to the primary policy. The evaluation of policies based on their categorical relationships is unique and untested in the Delphi policy.

Challenges to the Categorical Delphi include the experts' subjective interpretation of categorical classifications and the selection of an appropriate set of experienced experts. Where the addendum metrics are mapped to different sections of the primary policy, or if the relationship between the metrics of both policies is one-to-multiple, the methodology becomes more challenging. There is also a risk of achieving incomplete consensus among experts if their opinions diverge strongly, leading to fragmented results.

3. Methodology

3.1 Research Design

The CAA and CIF standards are collections of stipulations that are collections of quality metrics. This study focuses on a sample of 11 CAA metrics within Continuous Quality Enhancements stipulations as a purposive interest from the collection of stipulations in CAA regulations. The quality managers' perceptions of the mapping of 19 CIF metrics within the same stipulations as the 11 CAA metrics are of interest. The stipulation of interest is "Continuous Quality Enhancement". This study seeks to explore if the intents of the CIF metrics are pre-existing, partially existing, or are new ideas in the stipulations.

Table 1. Summary of the research design

Research Questions	Method	Data Source	Research Tool	Questionnaire output
RQ1 RQ2	Quantitative Qualitative	CAA and CIF	Categorical Delphi Technique	From Delphi rounds, indications of CIF: Fully captured, Partially captured, Not captured.
RQ3	Quantitative	Questionnaire	Conformance and Kendall's concordance	Likert Scale Rankings from an additional survey

Table 1 summarizes the methods applied to the research questions. Qualitative and quantitative methods were applied to study the first two questions, and a quantitative approach was adopted to address the third question. Two categories of experts participated in this study: Group 1, senior managers, and Group 2, program-level managers. Group 1 managers have significant knowledge of the institutional regulatory policies and Group 2 experts are responsible for policy implementation. The study was conducted within a business school in the UAE; however, the higher education institution was anonymized as agreed upon by the participants.

To investigate RQ1 and RQ2, the Categorical Delphi is implemented as seen in figure 1. The exploration phase is used to contact and select the participants. Predetermined questions are the primary measuring instruments (Cheung, 2021) in the questionnaires. They retrieve the experts' reflections of the primary and addendum metrics. Analysis of the collated responses provide overviews of the "consensus rates" expert evaluations on if the addendum metrics are wholly, partially, or non-existence in the primary metrics. To investigate RQ3, a post-Delphi round questionnaire is implemented rating their quality attributions and ranking them relative to each other. Measures of central tendency and other statistical tools were implemented to analyze the outcomes of the Likert ratings. Expert opinions and judgments are used in statistical inference and decision making through knowledge evaluation (O'Hagan, 2019).

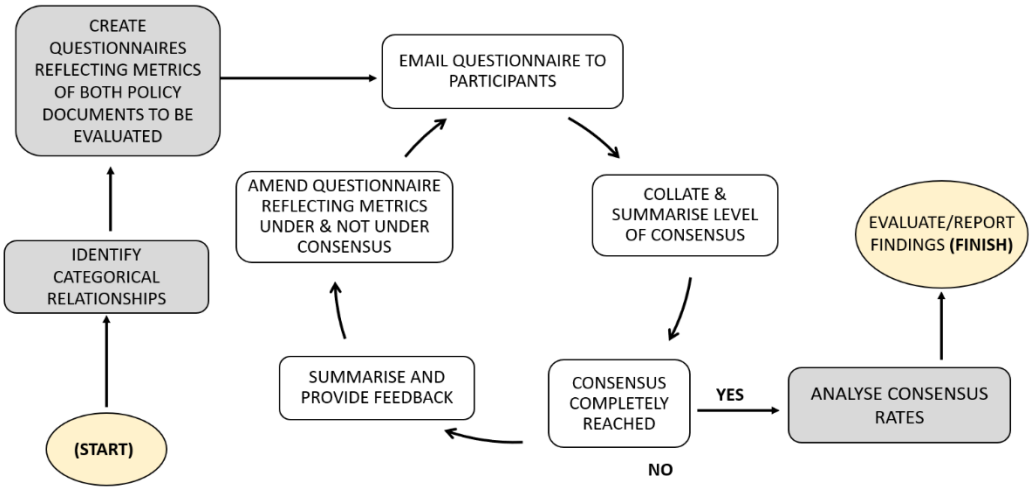

Figure 1. Protocol of the Categorical Delphi Technique (adapted from Couper, 1984)

3.2 Philosophical Underpinnings

The Delphi method is rooted in the philosophy of Locke, Kant and Hegel (Turoff, 1970). These philosophers stressed the significance of individual opinions and perceptions of the collective, along with other contributors of empirical data, in deconstructing the nature of reality, or ways of approaching decision-making. The Delphi technique hinges 'on the premise that "pooled intelligence" enhances individual judgement and captures the collective opinion of experts' (De Villiers et al., 2005. p.639). In this study, consensus is sought among participants following the Lockean inquiry system of truth as experimental, empirical with 'content that has been reduced to its observed behaviors, actions and impacts' (Manley, 2013, p.757). The nature of the Lockean inquiry system induces agreement and does not imbibe conflicts or debates when pertinent. This suggests that 'the inquiry system should be used only on a well-structured problem' (Mitroff & Turoff, 2002, p.22) with a 'strong consensual position on the nature of the problem situation' (p.22). This pragmatic research approach is philosophically consistent with Dewey's pragmatism. Dewey's pragmatism bridges the gap between theories and methods, stemming from the interpretive paradigm. Kirk and Reid (2002) noted that the Delphi technique capitalizes on subjective human experiences and context perspective.

3.3 Methods

A mixed quantitative and qualitative survey approach was implemented in the Categorical Delphi approach, and a strictly quantitative method was employed to determine Concordance and Kendall coefficients. The quantitative data involved ranking and rating quality statements derived from the CIF metrics. It also includes numerically counting the classifications of metrics in three categories: fully existing, partially existing, and non-existent within any CAA metric. The experts' critical reflections of the categorical relationships between the metrics of both policies are subjective evaluations and informed by their knowledge and experiences. Consensus rates were determined using a consensus scale of 0.0 – 1.0. The Policy Delphi was modified in this study and implemented for the purpose of mining expert opinions on the updates made to the primary policy's quality requirements using the addendum policy's metrics. The modifications are done to 'generate adequate information to sufficiently answer the research questions' (Creswell, 2013, p.4).

3.4 Data Collection and Analysis

In the first Delphi round, a questionnaire and the accompanying CIF and CAA spreadsheets were sent to Group 1 and Group 2 managers. The experts read through all the CIF and CAA metrics and evaluate whether the CIF metrics were fully captured, partially captured in the CAA metrics, or entirely different from the primary policy metrics. If the CIF metric is fully captured within the CAA metric, it is judged "Y". If the CIF metric is partially captured within the CAA, "P" is indicted next to the matched CAA metric. If the CIF is not captured within any of the CAA metrics, "N" is indicated as not captured. The responses of each group were collated.

In the second round of Delphi, summarized anonymized outcomes of the first Categorical Delphi round are shared with the managers within their clusters and not across clusters. The iterative nature of research exposes participants to the responses of other participants (Garavalia & Gredler, 2004). This is a positive attribute for improving the consensus. However, managers who are insecure with their first-round decisions may simply change their choices to

align with the majority. Managers are allowed to review and adjust their choices made in the first round of Delphi. At the end of two weeks of the second Categorical Delphi round, data were analyzed using the same approach employed at the end of the first round. The Delphi rounds were terminated at the end of the second round because of the achievement of 84.2% or more consensus by the two clusters of managers. The data generated includes metrics existing in the primary regulatory standard prior to the creation of the metrics of the policy addenda, partly existing in the CAA, and entirely new metrics not existing within the CAA. Where the CIF metrics are fully captured within the CAA, these metrics do not update the CAA; hence, they are not relevant for the research. Metrics not under consensus were also removed from the data owing to time limitations. All other CIF metrics are under consensus, either partially or entirely nonexistent within the CAA. These are classified as new metric additions by the CIF to the CAA regulations. The consensus rate was estimated as the proportion of participants in agreement. Based on the number of participants in each group of experts, the possible consensus rates were 0.0 (no consensus reached), 0.7 (2/3 consensus amongst experts) and 1.0 (3/3 consensus). Like Mercugliano (2025), 0.7 is adopted as the minimum measure of compliance.

To investigate the managers' view of the improvements made by the policy addendum on the primary policy, the degree of agreement between senior managers and middle managers was investigated using questionnaires beyond the Delphi rounds. The participants indicated their perception of the level of improvement made by addendum metrics to the primary metrics using a 5-point Likert rating. The measure of central tendency was then employed to describe expert opinions. Kendall's concordance (W) and chi-square values were also estimated. This is aimed at describing the level of agreement between the ratings of the experts within each group on the quality level of the newly created CIF metrics.

4. Results

4.1 Articulated Updates Made to the Primary Policy by the Addendum Policy Metrics

Perceptions of the categorical updates made by the addendum policy to the primary regulatory standard are captured from the consensus arrived at by the two clusters of managers.

Table 2. CAA and CIF metrics under consensus at the end of Delphi rounds 1 and 2

Group & Round	CAA Metrics Under Consensus		CIF Metrics Under Consensus	
Group 1 Round 1	1	5/11	S2.2.1.A, S2.2.1.B, S2.2.1.C, S2.2.1.D, S2.2.2.A.	12/19
Group 2 Round 1	2	6/11	S2.2.1.A, S2.2.1.B, S2.2.1.C, S2.2.1.D, S2.2.1.F, S2.2.1.G.	13/19
Group 1 Round 2	1	7/11	S2.2.1.A, S2.2.1.B, S2.2.1.C, S2.2.1.D, S2.2.1.F, S2.2.2.A, S2.2.3.	16/19
Group 2 Round 2	2	6/11	S2.2.1.A, S2.2.1.B, S2.2.1.D, S2.2.1.E, S2.2.1.F, S2.2.1.G.	17/19

Table 2 presents the consensus outcomes of the Categorical Delphi Analysis of the Group 2 experts. This shows that only 6/11 CAA metrics were updated by 13/19 and 17/19 CIF metrics at the end of Delphi rounds 1 and 2, respectively. Group 1 managers indicated that 5/11 and 7/11 CAA metrics were directly improved by 12/19 and 16/19 CIF metrics from the first to second Delphi rounds, respectively. The experts classified these metrics under consensus. The sum of the "partially captured" and "not captured" not captured' CIF metrics at the end of the Delphi rounds for the senior managers that the metrics under consensus rose from 63.2% to 84.3%. Likewise, the outcomes of the middle managers show that CIF metrics under consensus increased from 68.4% to 89.8% between the Delphi rounds. Overall, consensus describes agreements between managers that CIF metrics are fully or partially pre-existing in CAA metrics. There is also agreement between managers that some of the CIF metrics are entirely new additions to the quality requirements and are non-existent in the CAA metrics.

4.2 Relationships between the Primary and Addendum Policy Metrics

4.2.1 Between-Rounds Differences (G1R2 - G1R1) and (G2R2 - G2R1)

The "between round differences" provides indications of the rates of achievement of consensus between Delphi rounds. Differences were calculated between the Delphi rounds for each group of participants. Group 1 or senior managers' outcomes after the second Delphi round indicate that there is a 20.9% difference or increase in the CIF

metrics identified as fully or partially preexisting within the CAA metrics. However, the outcomes on column "D" also indicate that the CIF metrics identified as not pre-existing in the CAA metrics remained static at 6. This shows that there are no intersections between 31.6% of the CIF and CAA metrics.

Group 2 experts or middle managers' outcomes show that there was no change in the consensus rates at the end of the second Delphi round. Where the changes observed in the outcomes of Group 1 show that consensus is increasingly achieved between the opinions of the experts over Delphi iterations, no increases were observed in the Group 2 outcomes at the end of the second Delphi round. The outcomes of column (D) also show that there are no changes in the decisions of Group 1 experts on CIF metrics not pre-existing in the CAA. At the end of the second Delphi round, middle managers indicated that 57.9% of the CIF metrics were not fully or partially pre-existing in the CAA metrics. These metrics are classified as entirely new in the regulatory policy space. However, 31.6% of the CIF metrics were either partially or fully pre-existing in the CAA metrics.

4.2.2 Between-Groups Difference (G1R1 - G2R1) and (G1R2 - G2R2)

Table 3. Consensus at the end of the second Delphi round, between rounds, and within-group changes

	(A)	(B)	(C)	(D)	Between Rounds Difference		Between Group Difference	
	Fully pre-existing	Partly pre-existent	Not under Consensus	Non pre-existent	G1R2 - G1R1	G2R2 - G2R1	G1R1 - G2R1	G1R2 - G2R2
(G1R1)	0	6	7	6				
Group 1, Round 1	0.0%	31.6%	36.8%	31.6%				
(G1R2)	1	9	3	6				
Group 1, Round 2	5.3%	47.4%	15.8%	31.6%				
(G2R1)	4	2	6	7				
Group 2, Round 1	21.1%	10.5%	31.6%	36.8%				
(G2R2)	4	2	2	11				
Group 2, Round 2	21.1%	10.5%	10.5%	57.9%				
Change					20.9%	0.0%	0.0%	21.1%

This exhumed the difference in metrics under consensus between Group 1 and Group 2 experts at the end of each Delphi round. Table 3 shows that at the end of Delphi Round 1, there is no difference between the percentage of CIF metrics under consensus between middle and senior managers. At the end of the first Delphi round, the senior managers in Group 1 indicated that 31.6% of the CIF metrics partially pre-existed in the CAA metrics, while Group 2 or middle managers identified 21.1% of the CIF metrics as fully pre-existing in the CAA and 10.5% as partially pre-existing in the CAA. Differences were observed between the groups at the end of the second round of the Delphi. Where 52.7% of the CIF metrics are identified by Group 1 senior managers as fully or partially pre-existing in the CAA metrics, there are no changes in the positions of the middle managers. Hence, at the end of the second Delphi round, the between-group change in the CIF metrics that were fully or partially pre-existing in the CAA metrics was 21.1%. At the end of the second Delphi round, middle managers indicated that there were 26.3% more CIF metrics not pre-existing in the CAA metrics than were identified by senior managers.

4.3 Value Attributed by the Addendum Metrics to the Primary Policy

The sample sizes for rating each metric were three participants; hence, normal distribution does not apply, and other tests, such as the chi-square tests, were not applied. Estimations of central tendency were used to analyze the responses of Groups 1 and 2 experts. The differences in the list of newly created CIF metrics agreed upon by the groups were previously identified.

Table 4. Managers' ratings of quality attributes of CIF metrics

CIF Metric No.	Group 1 Experts		Group 2 Experts	
	*Mean	Value	*Mean	Value
C3.1.2	5.00	Excellent	4.67	Excellent
C3.1.3	3.67	Very good	4.33	Excellent
C3.2.3	4.00	Very good	x	x
C3.3.2	3.00	Good	x	x
C3.3.3	x	x	4.33	Excellent
C3.3.4	x	x	4.33	Excellent
C3.3.6	x	x	4.00	Very Good
C3.3.7	x	x	3.67	Very Good
C3.3.8	3.33	Good	3.33	Good
C3.3.9	x	x	4.00	Very Good
C3.3.10	3.33	Good	3.33	Good
C3.4.1	x	x	3.67	Very Good
C3.4.2	x	x	4.00	Very Good
*($0 \leq x \leq 1$)=Poor, ($1 < x \leq 2$)=Fair, ($2 < x \leq 3$)=Good, ($3 < x \leq 4$)=Very Good,($4 < x \leq 5$)=Excellent				

Table 4 presents the analysis of the experts' value attribution of the addendum policy metrics to the primary policy's Continuous Quality Enhancement metrics using a 5-point Likert scale. Group 1 and Group 2 experts agreed that the four metrics C3.1.2, C3.1.3, C3.3.8, and C3.3.10 were newly created, but they were not in concordance with the other identified metrics. The groups agree that C3.3.8, and C3.3.10 are good additions to the regulatory requirements. They also agreed that C3.1.2, is an excellent quality addition. C3.1.3 is perceived by Groups 1 and 2 experts as a "very good" and an "excellent" addition. Generally, senior and middle managers agree that the newly created CIF metrics are good, very good, or excellent additions to regulatory requirements.

Outcomes seen in Appendix 1 provides the within-group agreements in rankings of the "level of quality" the CIF newly created metrics bring to the regulations. Kendall's coefficient of concordance (W) of both groups moderately agrees that the experts concur with the strength of the contribution of these metrics to the regulatory requirements. The value of concordance at 0.0 and 1.0 indicates no agreement and complete agreement respectively. The experts also rated the metrics between 3 and 5, or good to excellent. The χ^2 values for both groups were less than the tabular values; hence, the null hypotheses were accepted. Group 1 and Group 2 rankings had a spread of 3-5 respectively with tied rankings.

The results provided in Appendix 2 present the outcomes of Kendall's concordance (W) and Chi-squared (χ^2) values of each group's ranking of the importance of the CIF addendum metrics compared with each other. Using a significance level of 0.05, W=0.54054, and χ^2 value=8.10812 for Group 1, which is below the expected value. This suggests a moderate agreement between Group 1 experts in their ranking of the newly created CIF metrics. The χ^2 value was also just below the expected value, thus agreeing with W. The χ^2 value is 16.79885 with a degree of freedom of 11 and a significance level of 0.05, and the null hypothesis is rejected, indicating agreement between the decisions of senior managers and middle managers.

5. Discussion

5.1 Articulated Updates Made to the Primary Policy by the Addendum Policy Metrics

The findings show that the consensus rates of Group 2 experts increased from 13/19 to 17/19 for the CIF metrics to the 6/11 CAA regulatory metrics. It thus reveals that the use of the Categorical Delphi approach distinctly isolated the relationships between the regulatory metrics and the addendum regulatory metrics. The revision of decisions made in the first Categorical Delphi round by the managers in the second round buttresses the position of Helmer (1967) that such revisions of initial judgements from the first Delphi round are sometimes revised by the participants. The change in position is likely due to reasons not captured in the data; however, such decisions result in

decision-makers taking a position. These positions are fundamental to how and whether the HEI effectively addresses the regulatory metric.

The Categorical Delphi Technique and its effectiveness in discerning the updates made by the addendum policy to the primary regulatory policy are highly subject to the effective definitions of categorical update types. Thus, the findings agree with Hsu and Sandford (2007) that poor categorical definitions of these types of updates could result in the omission of themes. The researcher is required to have significant knowledge of "categorical update types" within the policy context. Within HEIs, the Categorical Delphi would benefit from collective definitions of the perceived categorical update types.

Gordon and Pease (2006) indicated that the Delphi technique is research-intensive; it was found in this study to agree with Du Plessis (2007) that it is time demanding. Considering that only 11 CAA metrics and 19 CIF metrics were explored in this study, a considerably higher amount of time will be required to feasibly study the updates made by all metrics of the addendum policy on the primary policy. Turoff (1970) suggests that quantifying qualitative data can oversimplify nuances, especially in Policy Delphi. The findings in this study support the quantification of the outcomes of the qualitative categories, mainly because they are categorical, unambiguous, explicit, and direct. Categorical data type groups observations into distinct, qualitative categories without numerical significance (Agresti, 2018). Moore and McCabe (2012) viewed categorical data types as essential for summarizing group characteristics. In this instance, distinct categories are created of the "update types." Contrary to the categorical structure of the data retrieved in the "Categorical Delphi Technique," the data retrieved in Policy Delphi is simply qualitative (Linstone & Turoff, 1975) or quantitative (Dalkey & Helmer, 1963) and non-categorical.

5.2 Relationships between the Primary and Addendum Policy Metrics

Comparisons of Group 1 and 2 metrics under consensus show large differences between and within the positions of middle managers and senior managers. As shown in Table 3, the Categorical Delphi Technique outcomes revealed a tendency for the experts within their clusters to increasingly agree on the percentages of the primary metrics and addendum metrics under consensus. The outcomes of the Group 1 managers at the end of the second Delphi round show that only 31.6% of the CIF metrics are entirely new ideas in the policy space. At the end of Delphi round 2, the outcomes of Group 2 middle managers indicate that 57.9% of the metrics are new ideas in the policy space. The outcomes of Group 1 senior managers support Dalkey and Helmer's (1963) statement that progressive consensus is achieved by the Delphi Technique over Delphi rounds. However, no progressive improvements in consensus were observed in the outcomes of the group 2 managers. The use of the Categorical Delphi approach exposes deviations in consensus between middle managers and senior managers. These positional differences in the outcomes between the two groups of managers could reflect their positions on the policy implementation staircase (Reynolds and Saunders, 1987). Thus, the collective perceptions of the group of experts are consequential to the effectiveness of the Categorical Delphi Technique. The accuracy could be improved by improving expert selection, however, confining the group of experts to senior managers is managerial. The trajectory of the between-group expert consensus suggests that more iterations of the Delphi rounds would reduce the difference in consensus rates between senior and middle managers.

5.3 Value Attributed by the Addendum Metrics to the Policy Space

Generally, senior and middle managers agree that the newly created CIF metrics are good, very good, or excellent additions to regulatory requirements. Kendall's coefficient of concordance of both groups moderately agrees (Legendre, 2010). In this research, the experts concur on the strength of the attribution of these metrics to the primary regulatory requirements. The experts' rating of the strength of attribution of the addendum metrics between 3 and 5, or good to excellent, suggests that these metrics are viewed as positive contributions to the policy environment. The χ^2 values for both groups were less than the tabular values; hence, the null hypotheses were accepted. Rankings by the senior and middle managers of the strength of attribution of these metrics to the primary regulatory policy are thus spread between 3-5 and tied rankings.

Kendall's coefficient of concordance for the rankings of both groups was approximately 0.9, which is considered a "very good agreement" (Landis & Koch, 1977). Using a significance level of 0.05, Kendall's Coefficient of 0.54054 for Group 1, suggesting a moderate agreement between Group 1 experts in their rating of the importance of addendum metrics in relation to each other. Kendall's coefficient of 0.61505 indicates "good agreement" within Group 2 experts in their ranking of the newly created CIF metrics. The findings also indicate that there is agreement between the ratings of senior managers and middle managers of the importance of each of the addendum metrics relative to each other. Some addendum metrics were found to be more positive than others in the improvement of quality requirements.

6. Conclusion

The implementation of a Categorical Delphi Technique requires that the "update types" are categorical. The Categorical Delphi Technique is applied because regulatory metrics are itemized with singular ideas or requirements. Similar singular ideas are captured by the addendum policy metrics. Hence, the difference between the intents of the metrics of both policy documents can be deduced by experienced quality practitioners within the higher education sector.

The modification made to the Delphi technique was also able to identify the primary policy metrics recreated within the addendum policy. Categorical Delphi as described and implemented in this study categorically evaluates the relationships between two related regulatory policies.

There were differences in the achieved consensus and CIF metrics under consensus between senior and middle managers. A continuous improvement in the number of metrics under consensus was observed by one group of experts with increasing Delphi rounds. No changes were observed in the consensus rate of any group of experts. This suggests that it is not definitive that one or more Delphi rounds would be sufficient to conclude the study. The research outcome also shows an increasing definition of the number of CIF metrics that are not in consensus. Hence, it shows improved clarity regarding the number of CIF metrics that are newly introduced into the regulatory space.

Both groups of managers view the addendum metrics as good-to-excellent additions to the regulatory requirements. The similarity in the outcomes of the rankings of the CIF quality metrics by the two groups of managers suggest that their views of the addendum metrics are common.

7. Limitations

Where the number of regulatory metrics is very large, the analysis of data from implementing this approach would be complicated. The Categorical Delphi Technique is effective if a small number of highly knowledgeable policies are consulted. As suggested by Habibi et al. (2014), the robustness of the outcomes could be improved by increasing the number of experts.

The use of three experts draws from the concept of "an oracle." Although Rowe and Wright (2001) suggested that expert groups sizes should be between 5 and 20, selection of senior policymakers or middle managers for the study is purposive. Although the robustness of the evaluations would be enhanced by increasing the group size significantly above 3. Very large group sizes also introduce a degree of complexity to the study. However, irrespective of the number of participants, the findings would always be non-generalizable beyond the HEI environment or collection of HEIs.

References

Agresti, A. (2018). *An Introduction to Categorical Data Analysis (3rd ed.)*. Wiley.

Bollaert, L. (2014). A manual for internal quality assurance in higher education: With a special focus on professional higher education. *Eurashe*.

Broucker, B., & De Witt, K. (2015). "New Public Management in Higher Education." In *the Palgrave International Handbook of Higher Education Policy and Governance*. Edited by Jeroen Huisman, Harry de Boer, David D. Dill, and Manuel Souto-Otero, 57–75. New York: Palgrave Macmillan. https://doi.org/10.1007/978-1-137-45617-5_4

CAA, (2019). *Standard for Institutional Licensure and Programs Accreditation*. Ministry of Education, United Arab Emirates, 6th Edition.

Carroll, L. (2004). Clinical skills for nurses in medical assessment units. *Nursing Standard (through 2013), 18*(42), 33. https://doi.org/10.7748/ns.18.42.33.s56

Cheung, A. K. L. (2021). Structured questionnaires. In *Encyclopedia of quality of life and well-being research* (pp. 1-3). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-69909-7_2888-2

Cibulka, J. G. (1994). Policy analysis and the study of education, *Journal of Education Policy, 9*(506), 105-25. https://doi.org/10.1080/0268093940090511

CIF. (2025). *Framework for the Compliance Inspection of Higher Education Institutions*. Retrieved January 21, 2025, from https://www.moe.gov.ae/En/ImportantLinks/Documents/Higher%20Education%20Compliance%20Framework.pdf

Creswell, J. W. (2013). Steps in conducting a scholarly mixed-methods study. *DBER. Speaker Series*. Paper 48.

Couper, M. R. (1984). "The Delphi technique: characteristics and sequence model" *Advances in Nursing Science, 7*(1), 72-77. https://doi.org/10.1097/00012272-198410000-00008

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

De Villiers, M. R., De Villiers, P. J., & Kent, A. P. (2005). The Delphi technique in health sciences education research. *Medical teacher, 27*(7), 639-643. https://doi.org/10.1080/13611260500069947

Du Plessis, E., & Human, S. P. (2007). The art of the Delphi technique: highlighting its scientific merit. *Health SA Gesondheid, 12*(4), 13-24. https://doi.org/10.4102/hsag.v12i4.268

Easton, D. (1965). *A Framework for Political Analysis*, Eaglewood Cliffs, NJ: Prentice Hall.

Garavalia, L., & Gredler, M. (2004). Teaching evaluation through modeling: Using the Delphi technique to assess problems in academic programs. *American Journal of Evaluation, 25*(3), 375-380. https://doi.org/10.1177/109821400402500307

Gordon, T., & Pease, A. (2006). RT Delphi: An efficient, "round-less" almost real time Delphi method. *Technological Forecasting and Social Change, 73*(4), 321-333. https://doi.org/10.1016/j.techfore.2005.09.005

Habibi, A., Sarafrazi, A., & Izadyar, S. (2014). Delphi technique theoretical framework in qualitative research. *The International Journal of Engineering and Science, 3*(4), 8-13.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of advanced nursing, 32*(4), 1008-1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Helmer, O. (1967). *Systematic use of expert opinions*.

Hsu, C., & Sandford, B. A. (2007). "The Delphi Technique: Making Sense of Consensus", *Practical Assessment, Research, and Evaluation, 12*(1), 10. https://doi.org/10.7275/pdz9-th90

IIEP. (2018). Linking External and Internal Quality Assurance. *International Institute for Education Planning Policy Brief*. IQA and Higher Education N°4.

Keeney, S. (2010). The Delphi Technique. In: K. Gerrish, and A. Lacey, eds. 2010. *The Research Process in Nursing*. 6th ed. London: Wiley-Blackwell, pp. 227–236.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-74. https://doi.org/10.2307/2529310

Legendre, P. (2010). *Coefficient of concordance*. Pp. 164-169 in: Encyclopedia of Research Design, Vol. 1. N.J. Salkind, ed. SAGE Publications, Inc., Los Angeles.1776 pp. ISBN: 9781412961271.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method: Techniques and applications*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Mercugliano, E., Boshoff, M., Dissegna, A., Cerizza, A. F., Laner, L., Indovina, A. E., ... De Mori, B. (2025). Wildlife management and conservation in South Africa: informing legislative reform through expert consultation using the Policy Delphi methodology. *Frontiers in Veterinary Science, 12*, 1549222. https://doi.org/10.3389/fvets.2025.1549222

Manley, R. A. (2013). The Policy Delphi: a method for identifying intended and unintended consequences of educational policy. *Policy Futures in Education, 11*(6), 755-68. https://doi.org/10.2304/pfie.2013.11.6.755

Mitroff, I., & Turoff, M. (2002). Philosophical and Methodological Foundations of Delphi, in H. Linstone & M. Turoff (Eds) *The Delphi Method: techniques and applications*, pp. 17-34.

Moore, D. S., & McCabe, G. P. (2012). *Introduction to the Practice of Statistics (7th ed.)*. W.H. Freeman.

O'Hagan, A. (2019). Expert knowledge elicitation: subjective but scientific. *The American Statistician, 73*(sup1), 69-81. https://doi.org/10.1080/00031305.2018.1518265

Ozga, J. (2000). *Policy Research in Educational Settings: Contested Terrain*, Buckingham, UK: Open University Press.

Padró, F.F. (2009). Favoring standards-based over self-review external review processes in university QA schemes. *Journal of the World Universities Forum, 2*(6), 97-106. https://doi.org/10.18848/1835-2030/CGP/v02i06/56634

Powell, C. (2003). The Delphi technique: myths and realities. *Journal of advanced nursing, 41*(4), 376-382. https://doi.org/10.1046/j.1365-2648.2003.02537.x

Reynolds, J., & Saunders, M. (1987). *Teacher responses to curriculum policy: beyond the 'delivery' metaphor,* in J. Calderhead (ed.) Exploring Teachers' Thinking. London: Cassell.

Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London, UK: Routledge. https://doi.org/10.4324/9780203867396

Rowe G., & Wright G. (2001). Expert Opinions in Forecasting: The Role of the Delphi Technique. In: Armstrong J.S. (eds) *Principles of Forecasting. International Series in Operations Research & Management Science, 30*. Springer, Boston, MA. https://doi.org/10.1007/978-0-306-47630-3_7

Sackman, H. (1975). *Delphi Critique*. Lexington Books, Lexington, Massachusetts.

Santiago, P., Tremblay, K., Basri, E., & Arnal, E. (2008). *Tertiary Education for the Knowledge Society*. Volume 2 - Special Features: Equity, Innovation, Labour Market, Internationalisation. Paris: OECD, ISBN 978-92-64-04652-8.

Schlotter, M., Schwerdt, G., & Woessmann, L. (2009). *Methods for Causal Evaluation of Education Policies and Practices: An Econometric Toolbox*. EENEE Analytical Report to the European Commission, No. 5. Brussels: European Expert Network on Economics of Education.

Seyfried, M., & Pohlenz, P. (2018). Assessing quality assurance in higher education: quality managers' perceptions of effectiveness. *European Journal of Higher Education, 8*(3), 258-271. https://doi.org/10.1080/21568235.2018.1474777

Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). (2015). *Brussels, Belgium*.

Stensaker, B., & Maassen, P. (2020). Quality assurance in higher education: A global perspective. *Higher Education Research & Development, 39*(5), 921-935.

Turoff, M. (1970). Design of a policy Delphi. *Technological Forecasting and Social Change, 2*(2), 149-171. https://doi.org/10.1016/0040-1625(70)90161-7

Turoff, M., & Linstone, H. A. (2002). *The Delphi method-techniques and applications*.

Van Vught, F., & De Boer, H. (2015). "Governance Models and Policy Instruments In *the Palgrave International Handbook of Higher Education Policy and Governance*, edited by Jeroen Huisman, Harry de Boer, David D. Dill, and Manuel Souto-Otero, 38–56. New York: Palgrave Macmillan. https://doi.org/10.1007/978-1-137-45617-5_3

Appendix

Appendix 1. Estimations of Kendall's Concordance (W) and Chi-squared (χ^2) of the CIF quality attributions

(i) Test of the W statistics. H_o: The 3 experts are not concordant with one another. Kendall's W=0.55994 Accept H_o. (ii) Chi-squared test statistic. H_o: The 11 metrics are not concordant with one another. χ^2=8.39912, Accept H_o.				(i) Test of the W statistics. H_o: The 3 experts are not concordant with one another. Kendall's W=0.55996 Accept H_o. (ii) Chi-squared test statistic. Ho: The 11 metrics are not concordant with one another. χ^2=18.45161, Accept H_o.		
CIF Metric No.	Group 1 Expert Rankings*			Group 2 Expert Rankings*		
	G1E1	G1E2	G1E3	G2E1	G2E2	G2E3
C3.1.2	5	5	5	4	5	5
C3.1.3	3	4	4	4	4	5
C3.3.3	3	5	4	4	4	5
C3.3.4	3	3	3	4	4	5
C3.3.6	3	4	3	4	3	5
C3.3.7	3	4	3	3	3	4
C3.3.8	x	x	x	3	3	5
C3.3.9	x	x	x	4	3	5
C3.3.10	x	x	x	3	3	4
C3.4.1	x	x	x	4	3	4
C3.4.2	x	x	x	4	4	4
P=0.05 n=6 m=3 *Likert Scale: 1 - 5				P=0.05 n=11 m=3 *Likert Scale: 1 - 5		

Appendix 2. Kendall's Concordance (W) and Chi-squared (χ^2) values for the ranking of the addendum metrics

(i) Test of the W-statistics. H_o: The 3 experts are not concordant with one another. Kendall's W=0.54054, Accept H_o. (ii) Chi-squared test statistic. H_o: The 11 metrics are not concordant with one another. χ^2=8.10812, Accept H_o.				(i) Test of the W-statistics. H_o: The 3 experts are not concordant with one another. Kendall's W=0.61505 Accept H_o. (ii) Chi-squared test statistic. H_o: The 11 metrics are not concordant with one another. χ^2= 18.45161, Reject H_o.		
		Group 1 Expert Rankings*			Group 2 Expert Rankings*		
CIF No.	Metric	G1E1	G1E2	G1E3	G2E1	G2E2	G2E3
C3.1.2		7	7	7	6	6	7
C3.1.3		5	6	6	6	6	7
C3.3.3		4	7	3	6	5	7
C3.3.4		5	5	3	6	6	7
C3.3.6		5	6	2	6	6	6
C3.3.7		5	6	3	5	5	6
C3.3.8		x	x	x	5	5	6
C3.3.9		x	x	x	6	5	7
C3.3.10		x	x	x	5	5	6
C3.4.1		x	x	x	6	6	6
C3.4.2		x	x	x	6	6	6
P=0.05, n=6, m=3, *Likert Scale: 1 – 7					P=0.05, n=11, m=3, *Likert Scale: 1 - 7		