PriMera Scientific Engineering
Volume 3 Issue 1 July 2023
DOI: 10.56831/PSEN-03-062
ISSN: 2834-2550

PriMera Scientific Publications

Validation of a Satisfaction Survey using Cronbach's Alpha and Aiken's V: Improving Data Quality

Type: Review Article
Received: May 25, 2023
Published: June 29, 2023

Citation:
Jessica Uribe Navarrete. "Validation Of a Satisfaction Survey using Cronbach's Alpha and Aiken's V: Improving Data Quality". PriMera Scientific Engineering 3.1 (2023): 40-52.

Jessica Uribe Navarrete*

University of Salamanca, Chile

***Corresponding Author:** Jessica Uribe Navarrete, University of Salamanca, Chile.

Abstract

This article focuses on the importance of expert validation in the construction and evaluation of questionnaires, particularly in the context of satisfaction surveys. The article highlights that the involvement of experts provides an overview and helps understand the limitations of the coefficients used, such as Aiken's V and Cronbach's Alpha, as well as their proper interpretation.

The text mentions that expert validation can also provide additional information about the interpretation of results and key concepts in the evaluation of measurement instruments. Furthermore, it points out that expert validation can be useful in assessing the agreement among experts' ratings in an impact study.

Next, the importance of Aiken's V and Cronbach's Alpha in evaluating the validity of data obtained through questionnaires is emphasized. Aiken's V assesses the convergent and discriminant validity of a questionnaire, while Cronbach's Alpha evaluates the internal consistency of scale items.

It is highlighted that obtaining positive results in both coefficients enhances the validity of data by demonstrating good convergence, discrimination, and consistency in measuring the target construct.

In conclusion, the article emphasizes that expert validation and the use of coefficients such as Aiken's V and Cronbach's Alpha are fundamental in improving the validity of data in satisfaction surveys and other measurement instruments. These tools provide valuable information to ensure the accuracy and reliability of the results.

Keywords: Continuous education; Professional development; Satisfaction survey; Content validity; Reliability

Introduction

Understanding that, as Martín Cuadrado (2019, p. 187) states in his article "Continuing Education in the New Millennium," "Professional and personal development is related to a permanent attitude of inquiry, posing questions, and searching for solutions within the context of organizational development. This professional development often implies career advancement through the adoption of new

roles". Following this line, authors highlight a greater challenge in incorporating technological attributes, as indicated by Padilla and Juárez (2006, p. 11), "...In the current context, dominated by constant and rapid technological change, training in the company—as a means of generating human capital—plays a very important role in strengthening competitiveness".

One of the most relevant aspects in Continuous Education is being able to systematically evaluate student satisfaction to review their expectations and strive for continuous improvement. However, this evaluation must be consistent with the process and structured in a way that the data obtained provide information for analysis and the implementation of improvements. The results of a Satisfaction Survey will only be valid if the methodology employed includes reviewing the consistency and reliability of the instrument, as otherwise the data obtained could lead to distractions or disturbances that do not contribute to continuous improvement or sustainable analysis. Despite establishing this, it is not an easy task to have an evaluation tool that is adaptable and allows for the validation of the instrument through known methodologies, as it is a concept difficult to define and related to reliability.

As Pedrosa, Suárez-Álvarez, and García-Cueto (2014) indicate, "Currently, content validity is considered a necessary condition (although not sufficient) for interpreting test scores. Finally, the combination of both qualitative and quantitative methods is understood as the most comprehensive procedure when conducting an in-depth content validity study".

The methodology used in this process is based on two processes that were adapted to the instrument's proper application in a heterogeneous context and the need to solve problems with a systemic approach due to the nature of the training activities. This means that we necessarily had to seek a model that would provide a comprehensive view of the entire process (CIPP Model) and allow for the collection of information at specific milestones in an organized and practical manner.

Once we had a Satisfaction Survey that could be related to the complete activity, the next step was to validate it through methods based on expert judgment, and with these results, proceed to validate the reliability.

Therefore, the objective of this study was to adapt and validate reliability, understanding it as the consistency or stability of a measurement. According to Kerlinger and Lee (2002), "A technical definition of reliability that helps solve both theoretical and practical problems is one that investigates how much measurement error exists in a measuring instrument, considering both systematic and random variance".

In this study, the reliability coefficient linked to homogeneity or internal consistency was used, namely α (Alpha), proposed by Lee J. Cronbach (1916-2001) in 1951.

According to Quero (2010), "Firstly, it is worth noting, as has already been hinted, that the Cronbach's alpha coefficient 'naturally' expresses the degree to which items measure the same variable: homogeneity. Thus, its original utility is aimed at calculating the reliability of an instrument whose items or reagents constitute a single domain, that is, a single variable or trait".

Finally, as indicated by González and Pazmiño (2015), "Cronbach's alpha appears as a simple and reliable way to validate the construct of a scale and as a measure that quantifies the existing correlation among the items that compose it".

In the process of validating research, expert judgment emerges as a valuable method to verify the reliability of the obtained results. Experts, recognized for their experience and knowledge in study, provide an informed opinion that contributes to strengthening the credibility of the research. In this regard, we have incorporated Aiken's V, which provides additional information on interpreting the results, thus enhancing our understanding of reliability and validity concepts that are relevant for evaluating measurement instruments, including satisfaction questionnaires. This understanding is useful for comprehending the limitations of these coefficients and how they should be interpreted to improve the results objectively.

By incorporating expert judgment in the research validation, the robustness and quality of the results are strengthened. The informed opinion of these experts adds additional support to the findings, increasing confidence in the obtained data and the draw conclusions. In line with what Escobar-Pérez and Cuervo-Martínez (2008) indicate, this entails "an informed opinion from individuals

with a trajectory in the subject, recognized by others as qualified experts in the field, who can provide information, evidence, judgments, and evaluations". In conclusion, expert judgment plays a significant role in research validation. Their contribution as a source of informed and qualified opinion in study helps verify the reliability of the results and strengthens the overall credibility of the research. By considering the perspective of these experts, a critical and rigorous analysis is ensured, enhancing the quality of the research and its impact in the scientific field.

Investigation Methodology

According to Marszowski et al., it is necessary to improve people's competencies by promoting the idea of lifelong learning, whether through training or education. This is mainly driven by factors such as population aging, globalization, and the resulting competition in professions, as well as the constant changes in technological systems that require individuals to continuously learn in order for companies to be more productive.

Hence, the information gathered through the Satisfaction Survey is relevant for the teaching and learning system implemented in academia. Evaluating each aspect within the process is crucial to determine where to focus resources with greater certainty.

The research design was non-experimental and cross-sectional, using the "Expert Method" for assessment. The validation process included statistical review of Standard Deviation, Variance, Coefficient of Variation, as well as validation through Cronbach's Alpha, with values above 0.8 to ensure the reliability of the Likert scale. Additionally, Aiken's V was incorporated as a measure to validate the content of the instrument used (Satisfaction Survey) through expert opinions.

According to Cabrero and Llorente (2013), expert validation is relevant because "it can be very useful for validating diagnostic instruments or information collection instruments, and in combination with reliability strategies such as Cronbach's alpha, it can help in the construction of valid and reliable instruments".

From the evaluations sent to 14 experts, a sample of 6 received reports was obtained.

Total Reports Received	6	43%
Total Reports Sent	14	

Table 1: Total Reports Sent/Received.

The variables of the study were as follows:

- *Criterion variable*: Coherence, relevance, and clarity of the questions that will be part of the measurement of the "Satisfaction Survey" of Training Activities, based on the CIPP model.
- *Explanatory variables*: Ratings of the questionnaire items on a Likert scale.

The evaluation questionnaire for assessing the clarity, coherence, and relevance of a series of items grouped by dimensions is presented below:

1. Coherence: The item shows a logical and noticeable relationship with the measured dimension.
    1. The item has no logical relationship with the dimension.
    2. The item has a tangential relationship with the dimension.
    3. The item has a moderate relationship with the dimension.
    4. The item is fully related to the dimension.
2. Relevance: The item is essential for the dimension and must be present to measure it.
    1. The item can be eliminated without affecting the measurement of the dimension.
    2. The item has some relevance, but another item is including it in its measurement.
    3. The item is relatively important.

4. The item is very relevant and must be included.

3. Clarity: The item is easily understood (syntactically and semantically).
    1. The item is not clear.
    2. The item requires significant modifications in the use of words according to their meaning or their arrangement.
    3. A very specific modification of some of the terms in the item is required.
    4. The item is fully related to the measured dimension.

Regarding data analysis, exploratory techniques were applied. Initially, the distribution of the sample by question and its rating was studied, and subsequently, the obtained responses on the Likert scale (Coherence, Relevance, and Clarity) were analyzed using descriptive techniques.

Results

Sample Description

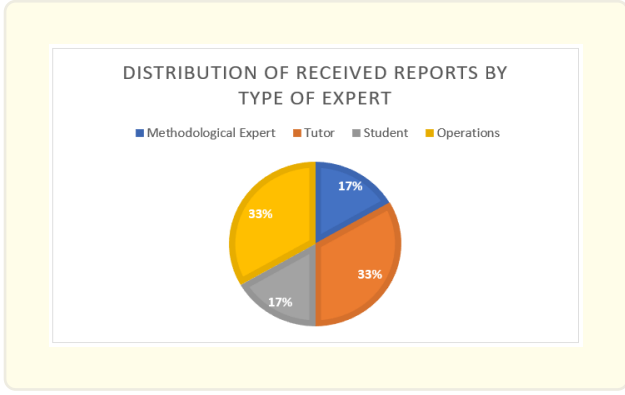
Expert validation is the most important part of this research because, although we rely on a validated model, the structure and questions must be coherent, understood by the public, and not biased towards one assessment or another. This ensures we have an objective instrument that truly serves to collect data for analysis and continuous process improvement.

In this case, three types of experts are considered, as shown in the following table:

Expert	***Definition***
Methodological Expert:	A professional with expertise in constructing data collection instruments, responsible for validating the standard structure and application of the model.
Subject Matter Expert:	A professional from the operations area of the Academy who provides support to students. Their knowledge is important as they are an integral part of the process and have a comprehensive understanding of it. Tutor: A professional responsible for mentoring and delivering knowledge to students during the execution of the activity. They have both theoretical and practical expertise in the content of the training program.
Student:	A participant who has completed training courses and is now providing feedback on the instrument. Their perspective is crucial in reviewing the instrument in terms of question format, length, and scope

Table 2: Expert Validation / Type of Expert.

As for their distribution, more than 60% of the participants are concentrated in the Tutor and Operations categories.



Descriptive Analysis of Ratings by Criteria

First, the reliability of the questionnaire consisting of 40 questions divided according to the CIPP model into Context (this item was not evaluated as it pertains to general knowledge of student expectations), Input (I), Process (P), and Product or Results (P) was analyzed.

Subsequently, the means, standard deviations, and variances of each question were obtained. These indicators were analyzed to assess their validity using the internal consistency coefficient, Cronbach's Alpha.

Based on this analysis, items with significantly higher variance (above 1.2) that had a negative impact on the index were eliminated. The instrument was considered valid when Cronbach's Alpha was greater than 0.7.

Dimension: "INPUT" Data

Firstly, Table 3 presents the distribution of ratings obtained for each question on the Likert scale, considering the criteria of Coherence, Relevance, and Clarity. It can be observed that the overall averages are above 3.0, except for questions 6 and 7. The dispersion (measured by the standard deviation) is generally low (not exceeding 1), except for questions 5, 6, 7, 8, and 9, indicating a high variance (above 1.5) for questions 5, 7, 8, and 9. In these cases, the ratings are not consistent.

Lastly, for these questions, the expert ratings show a distribution across multiple criteria ranging from 1 to 4, which indicates inconsistency. This suggests that these questions may not be well-developed and do not provide coherent information.

		Mode[a]	*Median*	*Mean*	*Standard Deviation*	*Coefficient of variation*	*Variance*	*Minimum*	*Maximum*
P1_1	Coherence	3.000	3.500	3.500	0.548	0.156	0.300	3.000	4.000
P1_1	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P1_1	Clarity	4.000	3.500	3.333	0.816	0.245	0.667	2.000	4.000
P2_1	Coherence	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000
P2_1	Relevance	4.000	4.000	3.500	0.837	0.239	0.700	2.000	4.000
P2_1	Clarity	4.000	4.000	3.500	0.837	0.239	0.700	2.000	4.000
P3_1	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P3_1	Relevance	4.000	4.000	3.500	0.837	0.239	0.700	2.000	4.000
P3_1	Clarity	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000
P4_1	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P4_1	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P4_1	Clarity	4.000	4.000	3.500	0.837	0.239	0.700	2.000	4.000
P5_1	Coherence	4.000	4.000	3.667	0.816	0.223	0.667	2.000	4.000
P5_1	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P5_1	Clarity	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P6_1	Coherence	1.000	1.500	1.833	1.169	0.638	1.367	1.000	4.000
P6_1	Relevance	1.000	1.500	1.833	1.169	0.638	1.367	1.000	4.000
P6_1	Clarity	3.000	3.000	2.333	1.033	0.443	1.067	1.000	3.000
P7_1	Coherence	3.000	3.000	2.500	1.225	0.490	1.500	1.000	4.000
P7_1	Relevance	3.000	3.000	2.500	1.225	0.490	1.500	1.000	4.000
P7_1	Clarity	4.000	3.500	3.167	1.169	0.369	1.367	1.000	4.000

P8_1	Coherence	4.000	4.000	3.167	1.329	0.420	1.767	1.000	4.000
P8_1	Relevance	4.000	4.000	3.000	1.549	0.516	2.400	1.000	4.000
P8_1	Clarity	4.000	4.000	3.333	1.211	0.363	1.467	1.000	4.000
P9_1	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P9_1	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P9_1	Clarity	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P10_1	Coherence	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000
P10_1	Relevance	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000
P10_1	Clarity	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000

[a] There is more than one mode, only the first one is reported.

Table 3: Distribution by Criteria and Question for Dimension 1.

Content Validity

When measuring the levels of validity and reliability using Cronbach's Alpha, the following can be established:

Scenario 1: Considering all the questions: If we observe Table 4, we can conclude that the levels of reliability and validity are not sufficient (below 0.5), except for clarity (above 0.8).

Cronbach's alpha Coherence	0,54
Cronbach's alpha Relevance	0,52
Cronbach's alpha Clarity	0,81

Table 4: Cronbach's Alpha with all the questions.

Scenario 2: When considering that questions 7, 8, and 9 generate significant distortion due to their high variance in all three criteria, with values above 1.5, and removing them from the survey, reliability and validity increase considerably:

Cronbach's alpha Coherence	0,91
Cronbach's alpha Relevance	0,84
Cronbach's alpha Clarity	0,92

Table 5: Cronbach's Alpha without questions 7, 8, and 9.

Dimension: "PROCESS (PROCES)" Data

Table 6 presents the distribution of ratings obtained for each question on the Likert scale for the criteria of Coherence, Relevance, and Clarity. It can be observed that the overall averages for all questions are above 3.0. The standard deviation is generally low (not exceeding 1), except for questions 1, 3, 4, 5, 8, and 9, indicating that the ratings for these questions are not homogeneous.

Furthermore, the expert ratings for these questions show a distribution across multiple criteria ranging from 1 or 2 to 4, which is not consistent. This suggests that these questions may not be well-developed and do not provide coherent information.

		Mode[a]	*Median*	*Mean*	*Standard Deviation*	*Coefficient of variation*	*Variance*	*Minimum*	*Maximum*
P1_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P1_2	Relevance	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P1_2	Clarity	2.000	3.000	3.000	0.894	0.298	0.800	2.000	4.000
P2_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P2_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P2_2	Clarity	4.000	4.000	3.833	0.408	0.106	0.167	3.000	4.000
P3_2	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P3_2	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P3_2	Clarity	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P4_2	Coherence	4.000	3.500	3.167	1.169	0.369	1.367	1.000	4.000
P4_2	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P4_2	Clarity	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P5_2	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P5_2	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P5_2	Clarity	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P6_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P6_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P6_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P7_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P7_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P7_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P8_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P8_2	Relevance	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P8_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P9_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P9_2	Relevance	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P9_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P10_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P10_2	Relevance	3.000	3.000	3.333	0.516	0.155	0.267	3.000	4.000
P10_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P11_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P11_2	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P11_2	Clarity	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P12_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P12_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P12_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P13_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P13_2	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P13_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000

P14_2	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P14_2	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P14_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P15_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P15_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P15_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P16_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P16_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P16_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P17_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P17_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P17_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P18_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P18_2	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P18_2	Clarity	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P19_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P19_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P19_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P20_2	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P20_2	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P20_2	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000

[a] There is more than one mode, only the first one is reported.

Table 6: Distribution by Criteria and Question for Dimension 2.

Content Validity

When measuring the levels of validity and reliability using Cronbach's Alpha, the following can be established:

Scenario 1: Considering all the questions, if we observe Table 7, we can indicate that the levels of reliability and validity are not sufficient (less than 0.5), except for Coherence (greater than 0.6, close to 0.7).

Cronbach's alpha Coherence	0,66
Cronbach's alpha Relevance	0,52
Cronbach's alpha Clarity	0,14

Table 7: Cronbach's Alpha with all the questions.

Scenario 2: When considering that questions 3, 4, and 5 generate significant distortion due to their high variance in all three criteria, with values above 1.3, and removing them from the survey, the reliability and validity increase considerably.

Cronbach's alpha Coherence	0,94
Cronbach's alpha Relevance	0,79
Cronbach's alpha Clarity	0,82

Table 8: Cronbach's Alpha without questions 3, 4, and 5.

Dimension: Data of "PRODUCT OR OUTCOME (PRODUCT)"

Table 9 shows the distribution of ratings obtained for each question on the Likert scale for the criteria of Coherence, Relevance, and Clarity. It can be observed that the averages are generally above 3.0, except for question 10. The dispersion (standard deviation) is generally low (not exceeding 1), except for questions 6, 7, 8, and 10, indicating that the ratings for these questions are not consistent.

Furthermore, the expert ratings for these questions show a distribution across multiple criteria ranging from 1 or 2 to 4, which indicates that these questions may not be well-developed and do not provide coherent information.

		Mode[a]	*Median*	*Mean*	*Standard Deviation*	*Coefficient of variation*	*Variance*	*Minimum*	*Maximum*
P1_3	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P1_3	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P1_3	Clarity	4.000	4.000	3.667	0.816	0.223	0.667	2.000	4.000
P2_3	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P2_3	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P2_3	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P3_3	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P3_3	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P3_3	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P4_3	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P4_3	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P4_3	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P5_3	Coherence	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P5_3	Relevance	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P5_3	Clarity	4.000	4.000	4.000	0.000	0.000	0.000	4.000	4.000
P6_3	Coherence	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P6_3	Relevance	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P6_3	Clarity	4.000	4.000	3.333	1.033	0.310	1.067	2.000	4.000
P7_3	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P7_3	Relevance	4.000	3.500	3.167	1.169	0.369	1.367	1.000	4.000
P7_3	Clarity	4.000	3.500	3.167	1.169	0.369	1.367	1.000	4.000
P8_3	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P8_3	Relevance	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P8_3	Clarity	4.000	3.500	3.167	1.169	0.369	1.367	1.000	4.000
P9_3	Coherence	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P9_3	Relevance	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P9_3	Clarity	4.000	4.000	3.667	0.516	0.141	0.267	3.000	4.000
P10_3	Coherence	4.000	4.000	3.500	1.225	0.350	1.500	1.000	4.000
P10_3	Relevance	3.000	3.000	2.833	0.983	0.347	0.967	1.000	4.000
P10_3	Clarity	2.000	2.500	2.500	1.049	0.420	1.100	1.000	4.000

[a] There is more than one mode, only the first one is reported.

Table 9: Distribution by Criteria and Question for Dimension 3.

Content Validity

When measuring the levels of validity and reliability using Cronbach's Alpha, the following can be established:

Scenario 1: Considering all the questions, if we observe Table 10, we can indicate that the levels of reliability and validity are not sufficient (less than 0.5).

Cronbach's alpha Coherence	0,48
Cronbach's alpha Relevance	0,53
Cronbach's alpha Clarity	0,32

Table 10: Cronbach's Alpha with all the questions.

Scenario 2: When considering that questions 6, 7, and 8 generate significant distortion due to their high variance across all three criteria, with values above 1.0, removing them from the survey significantly improves reliability and validity. Question 10 was not considered due to discrepancies across all three criteria.

Cronbach's alpha Coherence	0,92
Cronbach's alpha Relevance	0,97
Cronbach's alpha Clarity	0,90

Table 11: Cronbach's Alpha without questions 6, 7, and 8.

Finally, the questionnaire consists of 31 questions divided according to the following Table:

Dimension	***Number of questions***
Input (I)	7
Process (P)	14
Product or Result (P)	7

Table 12: Summary of Results - Number of Questions in Satisfaction Survey.

All items have an α above 0.79, ranging from 0.79 to 0.97.

Reliability of Expert Judgments

To assess the consistency or agreement of scores assigned by experts in the data analysis, the Aiken's V was used. This allowed us to evaluate the reliability of expert judgment by determining the correlation between the scores assigned by experts and the questions included in the "Satisfaction Survey".

The results obtained are presented in Table 13:

	Input (I)	***Process (P)***	***Product or Result (P)***
Coherence	0,23	0,28	0,29
Relevance	0,11	0,15	0,09
Clarity	0,17	0,16	0,18

Table 13: Aiken's V Correlation Results.

The previous results show a moderate Aiken's V for the "Coherence" of the questions in all three dimensions, indicating some agreement among the expert responses. This suggests a certain level of confidence in the quality of the evaluated questions, but there are still some differences in their opinions regarding the questions. However, when examining the criteria of relevance and clarity, the Aiken's V values range from 0.09 to 0.18, indicating a weak or very weak positive correlation between expert responses. This may suggest that significant changes are needed in the questions or the review process to improve the reliability of expert responses.

In summary, by combining the Aiken's V with Cronbach's Alpha, we can establish the following:

- The results obtained from the "Satisfaction Survey" in the INPUT dimension indicate that the coherence and relevance of the evaluated questions are moderate. The Cronbach's Alpha values for coherence and relevance are 0.54 and 0.52, respectively, while the Aiken's V values are 0.23 and 0.11, respectively. These values suggest that although there is some agreement among the evaluators' responses, there are still differences in their opinions regarding the coherence and relevance of the evaluated questions.
- On the other hand, the Cronbach's Alpha value for question clarity is high, at 0.81, indicating a high level of agreement among evaluators regarding the clarity of the questions. Additionally, the Aiken's V for clarity is very low, at 0.017, suggesting a high level of agreement among evaluators regarding the clarity of the evaluated questions.

Overall, these results indicate that the clarity of the evaluated questions is high, while coherence and relevance are moderate.

- The results obtained from the "Satisfaction Survey" in the PROCESS dimension indicate that the coherence and relevance of the evaluated questions are moderate. The Cronbach's Alpha for coherence is 0.66, and for relevance is 0.52. Additionally, the Aiken's V values for both dimensions are 0.28 and 0.15, respectively, suggesting a moderate positive correlation between the scores assigned by the evaluators in both dimensions.
- On the other hand, the Cronbach's Alpha value for question clarity is very low, at 0.14, indicating low coherence among evaluators' responses. Additionally, the Aiken's V for clarity is 0.16, suggesting a moderate positive correlation between the scores assigned by evaluators in this dimension.
- Overall, these results indicate that the coherence and relevance dimensions have a moderate correlation and are evaluated more consistently by the evaluators.
- The results obtained from the "Satisfaction Survey" in the PRODUCT dimension indicate that the coherence and relevance of the evaluated questions are low. The Cronbach's Alpha for coherence is 0.48, and for relevance is 0.53. Additionally, the Aiken's V values for both dimensions are 0.29 and 0.09, respectively, suggesting a moderate positive correlation for coherence and a weak positive correlation for relevance between the scores assigned by the evaluators.
- Furthermore, the Cronbach's Alpha value for question clarity is also low, at 0.32, indicating low coherence among evaluators' responses. Additionally, the Aiken's V for clarity is 0.18, suggesting a moderate positive correlation between the scores assigned by evaluators in this dimension.

Overall, these results indicate that all three dimensions of the survey have coherence and consistency issues among the evaluators. This may indicate that the questions are not well-designed or clear enough for the respondents. It is important to take steps to improve the quality of the survey to ensure that the obtained data is reliable and useful for decision-making.

Conclusions

Upon reviewing the data and information gathered to validate the "Satisfaction Survey," the following conclusions can be drawn:

1. The use of the CIPP Model provides an intuitive order that allows conducting the survey through consecutive dimensions, generating the complete process of the training activity. This enables a logical and easily interpretable sequence of the survey.
2. When examining the ratings, it is evident that they are mostly homogeneous despite the diverse perspectives, indicating that the survey has a coherent thread throughout.

3. By reviewing the Cronbach's Alpha and modifying (or eliminating) the questions that generated greater variability, we obtained a coefficient higher than 0.8 (0.79), validating the internal consistency of the survey and its reliability.
4. It is appropriate to make the indicated changes in the Satisfaction Survey to ensure a reliable instrument that provides important information for a comprehensive review of the training activity.
5. These results can be used to identify specific areas that require improvement in the satisfaction survey and, consequently, enhance the quality of the feedback provided by the respondents.

The results obtained for the instrument demonstrate consistency, a robust structure, adequate reliability, and validity. The methodology used allows for adaptation according to the specificity and relevance of the evaluation, without losing sight of the key points within the training activity, as it shows homogeneity in its different aspects.

Considering that reliability is a technical quality that involves accuracy and consistency in the collected information and is closely related to content validity (Pérez Juste et al., 2009), it is essential for research to have an acceptable level of rigor. In this regard, the Cronbach's Alpha coefficient is crucial in establishing this condition. Although it is a relatively straightforward procedure, it is methodologically robust.

It is important to understand that measuring the effectiveness of training is a way to achieve its potential impact and transfer educational interventions, leading to productive improvements in organizations and companies. It provides employees with more tools and motivation, which is relevant considering that both companies and the Chilean government invest significant resources in delivering training. Furthermore, as stated by Zineldin et al. (2011), we are aware that "student feedback on their experiences has become one of the central pillars of the quality process".

However, it is essential not to lose sight of the fact that this method has its limitations. Therefore, it is necessary to conduct a pilot project to further validate its implementation. In this regard, it is crucial to evaluate and ensure the final satisfaction of the users, emphasizing that this instrument is part of continuous improvement to deliver quality teaching and learning services, enabling the enhancement of competencies among the diverse collaborators in companies and organizations.

The validation of experts in questionnaire construction and evaluation is of great importance as it provides an overview and helps understand the limitations of the coefficients used and how to interpret them properly. Additionally, expert participation can provide additional information about result interpretation and key concepts in survey evaluation, especially in the case of satisfaction surveys. Furthermore, expert validation can also be useful in assessing agreement among experts in an impact study. In summary, expert validation plays a fundamental role in questionnaire construction, evaluation, and result comprehension, offering valuable insights and improving research quality.

The Aiken's V and Cronbach's Alpha are two widely used coefficients in evaluating the validity of data obtained through questionnaires. Although they have slightly different approaches and applications, both results are valuable for enhancing data validity.

Aiken's V is used to evaluate the convergent and discriminant validity of a questionnaire, examining to what extent the items on a scale are related to the construct they are supposed to measure and how distinct they are from other related constructs. It provides additional information for result interpretation, allowing a better understanding of whether the items effectively measure the intended construct. If the items demonstrate high convergent validity (significant correlation between them).

References

1. Cabrero Julio and Crying Carmen. "The application of expert judgment as a technique for evaluating information and communication technologies (ICT). Eduweb". Journal of Information and Communication Technology in Education 7 (2013): 11-22.
2. Escobar-Pérez J and Cuervo-Martínez A. "Content validity and expert judgment: an approach to its use". In Advances in Measurement 6 (2008): 27-36.

3. González Alonso J and Pazmiño Santacruz M. "Calculation and interpretation of Cronbach's Alpha for the case of validation of the internal consistency of a questionnaire, with two possible Likert-type scales". Publishing Magazine 2.1 (2015): 62-67.
4. Jornet Meliá., et al. The validity in the evaluation of programs (2000).
5. Martin Cuadrado AM. (s.f.). Impact of internal training on productivity and standardization of production processes: A case study.
6. Marszowski Ryszard., et al. "Education in the Times of Demographic Change and Globalization. Case Study on the Example of the Silesian Voivodeship". Sustainability 12.14 (2020): 5688.
7. Mingorance-Estrada AC., et al. "Validation of a questionnaire on the use of Interactive Response System in Higher Education". Rev. Latino-Am. Sick 29 (2021): e3418.
8. Padilla R and Juárez M. "Effects of training on the competitiveness of the manufacturing industry. ECLAC, ECLAC Subregional Headquarters in Mexico". Mexico D.F.: United Nations (2006).
9. Pedrosa Ignacio, Javier Suárez-Álvarez and Eduardo García-Cueto. "Content Validity Evidences: Theoretical Advances and Estimation Methods". Psychological Action 10.2 (2014): 3.
10. Quero Virla Milton. "Reliability and Cronbach's Alpha coefficient". Telos 12.2 (2010): 248-252.
11. Rodríguez-Rodríguez J and Reguant-Álvarez M. "Calculate the reliability of a questionnaire or scale using SPSS: Cronbach's alpha coefficient". REIRE Magazine of Innovation and Research in Education 13.2 (2020): 1-13.
12. Suárez Yerikson. Impact of Virtual Education and ICT on Andragogy (Educational Sciences Magazine, University of Carabobo) 28 (2018): 326-347.
13. Valencia Rodríguez WA and Vallejo Cardona JD. "Educational evaluation: more than an action, an ethical question". Virtual Magazine Catholic University of the North 45 (2015): 210-234.
14. Zineldin M, Akdag HC and Vasicheva V. "Assessing quality in higher education: new criteria for evaluating students' satisfaction". Quality in Higher Education 17.2 (2011): 231-243.