

Traducido con  DeepL

Reformulation of the K coefficient for expert judgement applied to digital teaching competencies

Reformulación del coeficiente k para juicio de personas expertas aplicada a las competencias digitales docentes

José González Campos^{*1}, Juan Elías Aspeé^{**}, Juan Silva Quiroz^{***} and Shadye Matar-Khalil^{****}

* Universidad Católica del Maule (Chile)

** Universidad Andrés Bello (Chile)

*** Universidad de Santiago de Chile (Chile)

**** Universidad del Sinú (Colombia)

Abstract

Expert judgment is a widely used technique to validate assessment, self-assessment, and self-perception instruments through informed item evaluation. The expert competence coefficient (K) combines self-assessment of knowledge and objective argumentation. It has been widely adopted in education and healthcare; however, its application as a simple, equally weighted average presents epistemological and statistical limitations due to its vulnerability to self-perception bias. A generalization of the K coefficient is proposed that incorporates a third component, referring to the expert's thematic proximity with differentiated weights for each dimension. An inferential model based on the beta distribution and the asymptotic normal approximation is integrated, allowing for the estimation of confidence intervals and significance tests on expert competence. It was applied in two contexts: a simulated case and a real-life validation study of an instrument on digital competencies in early childhood education. The analyses show that the traditional K coefficient overestimates competence compared to the generalized version. In the real-world case,

¹ **Correspondence:** José González Campos, jgonzalesc@ucm.cl. Facultad de Ciencias Básicas Universidad Católica del Maule

the sample initially classified as valid under the fixed threshold of 0.8 is reduced by applying the proposed inferential criteria, increasing the accuracy of expert identification. A test for paired samples confirms significant differences between the two methodologies. The reformulation of the K coefficient improves the rigor, transparency, and validity of expert judgment by considering both objective and contextual criteria and by supporting decisions through statistical inference. The proposal offers a robust and adaptable framework for future validation processes in various disciplines.

Keywords: evaluation; information and communication technologies; expert judgment; expert competence coefficient.

Resumen

El juicio de personas expertas es una técnica ampliamente utilizada para validar instrumentos de evaluación, autoevaluación, autopercepción, mediante la evaluación informada de ítems. El coeficiente de competencia experta (K) combina la autoevaluación del conocimiento y la argumentación objetiva, ha sido adoptado extensamente en educación y salud; pero su aplicación como promedio simple y equiponderado presenta limitaciones epistemológicas y estadísticas por su vulnerabilidad al sesgo de autopercepción. Se propone una generalización del coeficiente K que incorpora un tercer componente, referido a la proximidad temática del experto con ponderaciones diferenciadas para cada dimensión. Se integra un modelo inferencial basado en la distribución beta y la aproximación normal asintótica, permitiendo estimar intervalos de confianza y realizar pruebas de significación sobre la competencia experta. Se aplicó en dos contextos: un caso simulado y un estudio real de validación de un instrumento sobre competencias digitales en educación parvularia. Los análisis muestran que el coeficiente K tradicional sobrestima la competencia en comparación con la versión generalizada. En el caso real, la muestra inicialmente clasificada como válida bajo el umbral fijo de 0.8 se reduce al aplicar los criterios inferenciales propuestos, aumentando la precisión en la identificación de personas expertas. Una prueba para muestras pareadas confirma diferencias significativas entre ambas metodologías. La reformulación del coeficiente K mejora la rigurosidad, transparencia y validez del juicio experto al considerar tanto criterios objetivos como contextuales y al respaldar las decisiones mediante inferencia estadística. La propuesta ofrece un marco sólido y adaptable para futuros procesos de validación en diversas disciplinas.

Palabras Claves: evaluación; tecnologías de la información y la comunicación; juicio de expertos; coeficiente de competencia experta.

Introduction

Expert judgement is a systematic and rigorous procedure by which a group of individuals with a high degree of specialised knowledge evaluates, analyses and assesses aspects of an instrument, such as measurement instrument items, in order to determine its relevance, representativeness, coherence and clarity (Juárez-Hernández and Tobón,

2018). This methodological strategy has established itself as a fundamental resource in validation processes, because it allows for the integration of diverse and informed perspectives without requiring complex technical or logistical infrastructures. Thus, expert judgement contributes to the robustness and reliability of the evaluation process by ensuring the alignment of the instrument with the objectives and underlying theory of the study.

From a historical and epistemological perspective, one of the most significant advances in the objectification of expert judgement is the proposal of the coefficient of expert competence or K coefficient, initially proposed by Dobrov and Smirnov (1972). Over more than five decades, this coefficient has been widely adopted in various fields, especially in education and technology, as well as in other educational and health contexts. For example, there are some studies that show its versatility in the field of education with respect to entrepreneurial skills in students (Fernández Batanero and Reyes Rebollo, 2016), or in the perception of teaching teams with respect to inclusion (Torres and Fernández, 2015). In the field of health, it has had an impact on the selection of people as clinical tutors (Carrasco and Dois, 2020; Garrido-García et al., 2020) or on functional diversity (Aranda et al., 2022), rehabilitation (Capdevila-Pons et al., 2023), among others. Thus, its versatility is reflected in the diversity of applications, from the validation of instruments to the selection of professionals in clinical and educational settings.

In the methodological field, the literature has addressed aspects such as the optimal number of experts to guarantee the validity of the process, directly linked to the overcoming of minimum thresholds of the K coefficient (Herrera Masó et al., 2022). However, despite its widespread adoption, the metric structure itself and the statistical justification of the K coefficient are rarely subjected to critical analysis. In fact, an implicit acceptance of the original model, expressed through the formula $K = 1/2(Kc + Ka)$, persists without questioning either the logic of combining both components with equal weight, or the convenience of assigning alternative weights that better reflect their differentiated nature, an uncritical use that is common in the field of applied statistics (Gelman and Hennig, 2017).

Moreover, Dobrov and Smirnov (1972) themselves noted that the definition of the coefficient is susceptible to different configurations, depending on how its Kc (subjective self-assessment) and Ka (objective assessment based on external criteria) components are conceptualised and operationalised. The original proposal of calculating an equal-weighted arithmetic mean does not exclude the possibility of assigning differentiated weights, an idea that was taken up and further developed by Cabero-Almenara et al. (2020), who stress the need to establish explicit criteria for such weighting. Furthermore, Zарtha Sossa et al. (2014) raise concerns regarding the impact of sample size on the stability and representativeness of the coefficient, an aspect that has so far been insufficiently explored.

The operational and quantitative expression of the k-expert statistic is defined as $\frac{1}{2}(K_c + K_a)$, i.e. it is an average between the scores obtained in K_c and K_a , in which K_c represents the personal assessment, i.e. the potential expert's assessment of him/herself in relation to what is to be measured, while K_a represents the assessment in terms of criteria defined by the researchers who seek a particular profile of the experts who will constitute the sample of validators, respectively.

From the epistemological perspective or the genesis of the statistic, it is not made explicit whether it is intentional or coincidental, but it is ultimately a mediation between subjectivity (K_c) and objectivity (K_a), both of which have the same weight.

From the metric perspective, some inconsistencies and inaccuracies arise, as K_c and K_a have different metric statuses, which are corrected when obtaining K, and there is also doubt as to whether the weights for each component are equal and why the decision criterion is based on the constant 0.8 (Cerero et al., 2023).

Thus, a central problem that motivates this study is the presence of epistemological and methodological limitations in the current configuration of the K coefficient. This coefficient, by averaging two components of a heterogeneous nature (the subjective assessment of knowledge itself - K_c - and the objective assessment based on criteria defined by researchers - K_a -), can introduce biases if both elements are considered with the same weight without an adequate metric or theoretical justification.

This is important, since the reliability of the subjective K_c component is compromised by its very nature, according to the expert consulted. For example, there is the Dunning-Kruger effect, a widely documented phenomenon, which indicates that individuals with lower ability tend to overestimate their abilities, while the most competent tend to underestimate them (Kruger and Dunning, 1999; Mazor and Fleming, 2021; Vílchez, 2020). This overestimation or underestimation directly affects the accuracy of the coefficient, as it assumes as valid a self-assessment that could be biased. Likewise, the absence of an inferential framework that statistically supports the decision thresholds, such as the widespread use of the value 0.8 to define whether a person is considered an expert, limits the validity of the process, as there is a lack of a solid empirical basis to justify this choice (Gelman and Hennig, 2017). Therefore, the literature points to the importance of incorporating statistical methods that allow estimating confidence intervals and performing significance tests, thus providing more robust decision criteria (Barroso and Cabero, 2013; Cabero-Almenara et al., 2020; Dobrov and Smirnov, 1972).

This evidence suggests the need to review and reformulate the K coefficient, integrating differentiated weights and inferential criteria that improve the selection and validation of experts, particularly in educational contexts.

Faced with these limitations, this article proposes a generalisation and deepening of the metric configuration of the coefficient of expert competence, incorporating differentiated weights that reflect the different natures of the components and a robust inferential model based on the beta distribution and the asymptotic normal approximation, which allows for the establishment of well-founded statistical criteria for

the selection of experts. This reformulation seeks to provide the validation process with greater rigour, flexibility and objectivity, helping to overcome the epistemological and methodological deficiencies detected, especially in the context of the validation of assessment instruments. With this framework, it is valid to indicate that this paper will not discuss or question the conceptual complexity referred to the term expert, as it is beyond the scope proposed, however, it is possible to find interesting discussions in Ramírez and Cepena (2020), and in Juárez-Hernández and Tobón (2018).

Method

Type of research

The present study is framed within a quantitative research of a methodological and cross-sectional nature, oriented towards the analysis, reformulation and inferential validation of the coefficient of expert competence (K) in expert judgement processes. Its purpose is not only descriptive, but also incorporates a component of statistical modelling and empirical contrast between different configurations of the statistic.

The design contemplates two different applications of the expert assessment instrument: Application through controlled statistical simulations, in order to analyse the behaviour of the coefficient in theoretical scenarios under different weighting and distribution schemes. Application with real data from a validation process of an instrument on digital competences in kindergarten education (Silva-Quiroz et al., 2025), with the purpose of empirically contrasting the classical model and the generalised proposal.

This double approach, simulated and empirical, allows us to evaluate both the mathematical consistency of the model and its behaviour in real validation contexts.

Population and sample

Simulated case: In the simulated application, an artificial sample of 50 potential experts was generated, each characterised by three components:

K_c : knowledge coefficient (self-assessment),

K_a objective argumentation coefficient,

K_b thematic proximity coefficient.

The values were generated under probabilistic assumptions compatible with variables bounded in the interval [0,1], allowing us to model the behaviour of the generalised coefficient under different weighting schemes and to analyse its inferential stability.

Real case: In the empirical application, we worked with a sample of 32 potential experts who participated in the validation process of a digital competences instrument adapted to the context of kindergarten education in Chile. The original data included the classical dimensions K_c and K_a , from which the conventional coefficient was reconstructed and the proposed reformulation was applied.

Results

Formality and limitations of the K coefficient

Equating the components K_c and K_a , means to value equally subjectivity and objectivity, which is represented in $\frac{1}{2}(K_c + K_a)$. However, this model and the intention of its definition, as a criterion for the selection of experts, generates some epistemological contradictions referring to the Dunning-Kruger effect (Kruger and Dunning, 1999), a concept that is very present and strongly discussed in various fields, particularly in the educational environment (Estrada Girón et al., 2022; Vílchez, 2020), aspects that lead to the conclusion that subjectivity is overestimated in the specification of the K statistic, justifying a differentiated weighting in its generalisation.

Generalisation and extension: coefficient K_{eg}

Considering the above, it is essential to add a third component to the definition of the K coefficient that quantifies the degree of proximity of the expert with respect to the subject matter of the instrument and the type of knowledge that the expert possesses about it. Indeed, giving an opinion from theory deserves a differentiation in relation to opinion from practice, for this reason the generalisation and extension of the statistic K , which will be symbolised by K_{eg} is defined as:

$K_{eg} = \alpha K_c + \beta K_a + \gamma K_b$ where $\alpha + \beta + \gamma = 1$ and K_b represent the degree of proximity of the potential expert to the person to be evaluated or measured.

It is observed that when considering $\alpha = 1/2$ $\beta = 1/2$ and $\gamma = 0$, the conventional k -expert coefficient is obtained, placing it as a particular case.

Inferential process for expert decision

In order to introduce the concept of significance in the coefficient K_{eg} , it is necessary to formalise an inferential process around its definition. Therefore, it will be said that $K_{eg_1}, K_{eg_2}, K_{eg_3}, \dots, K_{eg_n}$ represents a sequence of independent and identically distributed random variables. In relation to the identical distribution, the K_{eg} will be assumed to have a beta distribution, of parameters p and q , which is symbolised by $B(p, q)$, in which the mean is $\frac{p}{p+q}$ and its finite variance is $\frac{pq}{(p+q)^2(p+q+1)} < \infty$. In this context we define $\bar{K}_{eg} = \frac{1}{n} \sum_{i=1}^n K_{eg_i}$, which, by construction and the central limit theorem, is $\bar{K}_{eg} \sim N(E(\bar{K}_{eg}); Var(\bar{K}_{eg}))$, where $E(\bar{K}_{eg}) = \frac{p}{p+q}$ $Var(\bar{K}_{eg}) = \frac{pq}{n(p+q)^2(p+q+1)}$

On the other hand, considering the sample estimates of the unknown parameters p and q , this is

$$\begin{aligned} \tilde{p} &= \bar{x} \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) \\ \tilde{q} &= (1-\bar{x}) \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) \end{aligned}$$

Where \bar{x} and s^2 represent the sample mean and variance respectively (González et al., 2014), the confidence interval for $\frac{p}{p+q}$ will be given by:

$$IC\left(\frac{p}{p+q}\right) = \left[\frac{\tilde{p}}{\tilde{p} + \tilde{q}} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{4\tilde{p}\tilde{q}}{n(\tilde{p} + \tilde{q})^2(\tilde{p} + \tilde{q} + 1)}}; \frac{\tilde{p}}{\tilde{p} + \tilde{q}} + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{4\tilde{p}\tilde{q}}{n(\tilde{p} + \tilde{q})^2(\tilde{p} + \tilde{q} + 1)}} \right]$$

Where the number 4 is a correction constant and α the significance level. Based on the above definition, it is possible to establish two decision criteria for the sample of experts.

In order to establish clear and well-founded rules for identifying experts within a sample, two complementary decision criteria are proposed: a stochastic one, based on statistical inference, and a deterministic one, based on a previously established conventional threshold.

a) Stochastic criterion: the analysis sample will be said to constitute a valid sample of experts if $1 \in IC\left(\frac{p}{p+q}\right)$. This perspective makes it possible to specify the set of specialists who can be considered in the category of expert person for the validation process, under the criterion of belonging or not to $IC\left(\frac{p}{p+q}\right)$. That is, if $K_e g_i$ represents the generalised expert index k of person i , and if $K_e g_i \in IC\left(\frac{p}{p+q}\right) \Rightarrow$ la i – ésim a person is considered an expert.

b) Deterministic criterion: Now if $1 \notin IC\left(\frac{p}{p+q}\right)$, then the conventional deterministic criterion can be used, i.e. the i – ésim a person can be considered an expert if $.K_e g_i \geq 0.8$

It is necessary to specify that the proposed stochastic criterion, based on the comparison of the confidence interval with the threshold value, should not be interpreted as a deterministic test of absolute membership, but as an indicative inferential decision rule. The presence or absence of the reference value within the interval reflects the statistical evidence available under the assumptions of the model, but does not imply an ontological or categorical classification of expert status. In this sense, the approach adopted seeks to introduce a probabilistic framework complementary to the conventional criterion, recognising that the decisions derived are conditioned by the sample size, the observed variability and the distributional assumptions made.

Explaining that the sample is valid for expert judgement means that, regardless of the existence of persons whose generalised k -expert coefficients are less than 0.8, the sample is sufficiently robust to validate persons for expert judgement.

Finally, when analysing the internal structure of K_a , based on its original source, Dobrov and Smirnov (1972), because all the other uses are respectful of this structure

referred to the organisation of the scores, which are presented in Table 1 (the original) below:

Table 1

Evaluation of the sources of argumentation

Source	Extent to which your answer was influenced by the source		
	H (High)	A (Average)	L (Low)
Theoretical analysis you have conducted [Theoretical analysis you have conducted].	0.30	0.20	0.10
Your practical experience	0.50	0.40	0.20
Generalization of works of native authors [Generalization of works of native authors]	0.05	0.05	0.05
0.05 0.05 0.05 0.05 0.05 0.05 Generalization of works of foreign authors	0.05	0.05	0.05
Your own knowledge of the world-level state of the art [Your own knowledge of the world-level state of the art]	0.05	0.05	0.05
Your intuition	0.05	0.05	0.05

Note. Source: Dobrov and Smirnov (1972). Own adaptation.

Table 1 above shows that the last four characteristics in the specification of the statistic K_c , the Low, Medium and High levels, are indistinguishable in score, therefore, the conclusions based on this same organisation of the scores is subject to the first two criteria. Therefore, the question arises as to what is the sense of defining 4 descriptors, whose answers do not generate any additional information to the expert's selection process. Based on this non-parsimonious specification, a significance comparison of the effect of this indifferentiation on the conclusions drawn is carried out.

Calculation of the weights

To establish the optimal weights in the definition of the generalised coefficient K_{eg} , the following method is proposed: three columns of data corresponding to K_c , K_a y K_b are considered, with n observations in each column. The objective then is to determine the weights α , β and γ that allow us to construct a weighted sum of the form $\alpha K_c + \beta K_a + \gamma K_b$ that maximises this expression under the assumptions that $\alpha + \beta + \gamma = 1$ and that all the weights are greater than zero ($\alpha, \beta, \gamma > 0$).

To obtain the weights α, β y γ , we follow these steps:

First, we calculate the averages of each column of data.

$$\begin{aligned}\bar{k}_a &= \frac{1}{n} \sum_{i=1}^n k_{ai} \\ \bar{k}_b &= \frac{1}{n} \sum_{i=1}^n k_{bi} \\ \bar{k}_c &= \frac{1}{n} \sum_{i=1}^n k_{ci}\end{aligned}$$

Standardisation of the averages. Let $S_p = \bar{k}_a + \bar{k}_b + \bar{k}_c$ be the sum of the averages. Then, the weights are determined as:

$$\begin{aligned}\alpha &= \frac{\bar{k}_c}{S_p} \\ \beta &= \frac{\bar{k}_a}{S_p} \\ \gamma &= \frac{\bar{k}_b}{S_p}\end{aligned}$$

We then use the weights obtained to calculate the weighted sum for each observation.

$$S_i = \alpha k_{ci} + \beta k_{ai} + \gamma k_{bi}$$

This section presents the results obtained from the application of the methodological proposal developed for the validation of experts in the context of a national research project aimed at assessing the level of digital competence of students in the final years of kindergarten education. The performance of a sample of potential experts is analysed using both the classical coefficient of competence (*K*) and the new proposed approach, which incorporates methodological adjustments aimed at strengthening the rigor of the selection process.

Application in a simulated case

Suppose that there are 50 potential experts to whom the generalised *k*-expert coefficient will be applied to assess their suitability in the calibration of any instrument.

Consider the estimation statistic of the generalised *k*-expert coefficient, $K_{eg} = \alpha K_c + \beta K_a + \gamma K_b$, with $\alpha = \beta = \gamma = \frac{1}{3}$. The estimates are described in Table 2.

Thus, one has that: $\bar{x} = 0.896, s^2 = 0.072$, then $\tilde{p} = \bar{x} \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) = 0.249 \tilde{q} = (1 - \bar{x}) \left(\frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right) = 0.028$, from which one has:

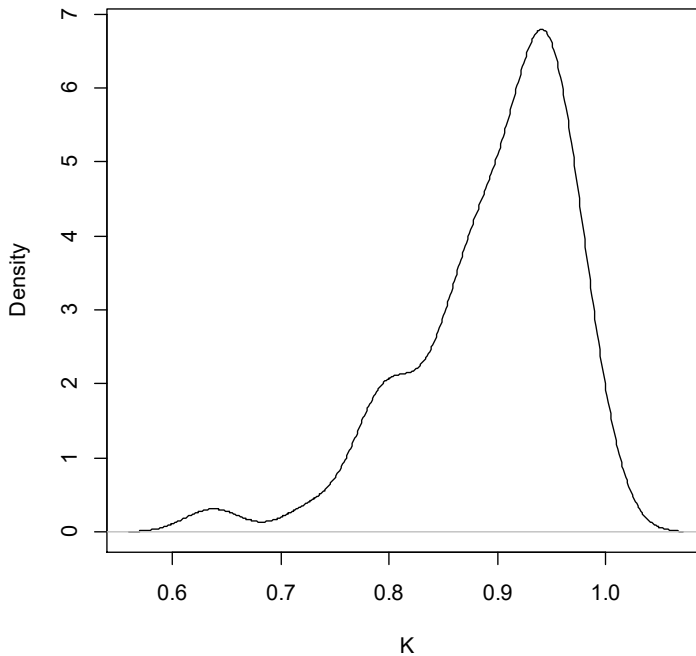
$$\begin{aligned}IC\left(\frac{p}{p+q}\right) &= \left[\frac{\tilde{p}}{\tilde{p} + \tilde{q}} - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{4\tilde{p}\tilde{q}}{n(\tilde{p} + \tilde{q})^2(\tilde{p} + \tilde{q} + 1)}}; \frac{\tilde{p}}{\tilde{p} + \tilde{q}} + z_{1-\frac{\alpha}{2}} \right. \\ &\quad \left. \cdot \sqrt{\frac{4\tilde{p}\tilde{q}}{n(\tilde{p} + \tilde{q})^2(\tilde{p} + \tilde{q} + 1)}} \right] = [0.746; 1.000]\end{aligned}$$

Then the analysis sample constitutes a valid sample of experts, because $1 \in IC\left(\frac{p}{p+q}\right)$. This means that the sample of potential experts will fully identify the validating experts.

It is also possible to plot the estimated distribution function for K_{eg} in Figure 1 below, from which a marked skewness to the left can be seen, implying a tendency towards high values, except for an outlier situation between 0.6 and 0.7, which gives support and consistency to the preliminary decisions.

Figure 1

Distribution of generalised K-expert in 50 simulated cases.



In contrast, from the classical perspective, considering $K = 1/2(Kc + Ka)$, and using the decision threshold of 0.8, the results presented in Table 2 show that 7 participants were not classified as experts for the validation process.

On the other hand, if we consider a lower measure of subjectivity, for example, $\alpha=0.3$, and consequently $\beta=0.7$, we obtain the results presented in Table 3. This differentiation in weighting, by giving greater relevance to the objective selection criteria, results in an increase in the number of participants considered ineligible for the validation process, leaving out only nine people who do not reach the established threshold, highlighting the relevance of the weightings.

Table 2

Simulated estimates for conventional generalised generalised *K*-expert

K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}
0,81	0,98	0,93	0,90	0,6	0,9	0,95	0,73	0,79	0,63	0,95	0,71	0,78	0,93	0,97	0,86	0,84	0,92	0,96	0,88
0,94	0,97	0,99	0,96	0,84	0,96	0,84	0,90	0,97	0,85	0,99	0,91	0,99	0,98	0,69	0,99	0,9	0,85	0,99	0,88
0,76	0,99	0,61	0,88	0,98	0,85	0,99	0,92	0,98	0,99	0,91	0,99	0,95	0,89	0,98	0,92	0,93	0,71	0,93	0,82
0,83	0,99	0,81	0,91	0,86	0,84	0,83	0,85	0,98	0,83	0,84	0,91	0,73	0,94	0,99	0,84	0,95	0,99	0,98	0,97
0,92	0,89	0,99	0,91	0,89	0,73	0,86	0,81	0,94	0,83	0,86	0,89	0,99	0,72	0,93	0,86	0,93	0,96	0,99	0,95
0,94	0,96	0,68	0,95	0,99	0,93	0,99	0,96	0,91	0,99	0,95	0,95	0,99	0,99	0,88	0,99	0,9	0,99	0,79	0,95
0,97	0,96	0,91	0,97	0,89	0,95	0,99	0,92	0,95	0,99	0,99	0,97	0,99	0,91	0,86	0,95	0,98	0,56	0,86	0,77
0,5	0,7	0,96	0,60	0,97	0,88	0,98	0,93	0,67	0,52	0,61	0,60	0,76	0,98	0,99	0,87	0,96	0,87	0,8	0,92
0,86	0,98	0,99	0,92	0,99	0,96	0,93	0,98	0,92	0,69	0,99	0,81	0,74	0,64	0,88	0,69	0,93	0,87	0,92	0,90
0,86	0,99	0,98	0,93	0,99	0,93	0,69	0,96	0,71	0,98	0,93	0,85	0,98	0,79	0,99	0,89	0,89	0,69	0,86	0,79

Note. Each row of Table 2 represents a simulated case, composed of its respective values for the three components of the generalised *K*-expert coefficient: the objective dimension (*K_c*), the subjective dimension (*K_a*) and the contextual dimension (*K_b*), considering $\gamma = 0$

Table 3

Simulated estimates for generalised K-expert with threshold 0.8

K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}	K_c	K_a	K_b	K_{eg}
0,81	0,98	0,93	0,93	0,6	0,9	0,95	0,80	0,79	0,63	0,95	0,68	0,78	0,93	0,97	0,89	0,84	0,92	0,96	0,90
0,94	0,97	0,99	0,96	0,84	0,96	0,84	0,92	0,97	0,85	0,99	0,89	0,99	0,98	0,69	0,98	0,9	0,85	0,99	0,87
0,76	0,99	0,61	0,92	0,98	0,85	0,99	0,89	0,98	0,99	0,91	0,99	0,95	0,89	0,98	0,91	0,93	0,71	0,93	0,78
0,83	0,99	0,81	0,94	0,86	0,84	0,83	0,85	0,98	0,83	0,84	0,88	0,73	0,94	0,99	0,88	0,95	0,99	0,98	0,98
0,92	0,89	0,99	0,90	0,89	0,73	0,86	0,78	0,94	0,83	0,86	0,86	0,99	0,72	0,93	0,80	0,93	0,96	0,99	0,95
0,94	0,96	0,68	0,95	0,99	0,93	0,99	0,95	0,91	0,99	0,95	0,97	0,99	0,99	0,88	0,99	0,9	0,99	0,79	0,96
0,97	0,96	0,91	0,96	0,89	0,95	0,99	0,93	0,95	0,99	0,99	0,98	0,99	0,91	0,86	0,93	0,98	0,56	0,86	0,69
0,5	0,7	0,96	0,64	0,97	0,88	0,98	0,91	0,67	0,52	0,61	0,57	0,76	0,98	0,99	0,91	0,96	0,87	0,8	0,90
0,86	0,98	0,99	0,94	0,99	0,96	0,93	0,97	0,92	0,69	0,99	0,76	0,74	0,64	0,88	0,67	0,93	0,87	0,92	0,89
0,86	0,99	0,98	0,95	0,99	0,93	0,69	0,95	0,71	0,98	0,93	0,90	0,98	0,79	0,99	0,85	0,89	0,69	0,86	0,75

Note. Each row of Table 3 represents a simulated case, composed of its respective values for the three components of the generalised K-expert coefficient: the objective dimension (K_c), the subjective dimension (K_a), considering $\gamma = 0$

Table 4

Simulated estimates for generalised equally weighted generalised K-expert

K _c	K _a	K _b	K _{eg}	K _c	K _a	K _b	K _{eg}	K _c	K _a	K _b	K _{eg}	K _c	K _a	K _b	K _{eg}	K _c	K _a	K _b	K _{eg}
0,81	0,98	0,93	0,91	0,6	0,9	0,95	0,80	0,79	0,63	0,95	0,79	0,78	0,93	0,97	0,89	0,84	0,92	0,96	0,91
0,94	0,97	0,99	0,97	0,84	0,96	0,84	0,88	0,97	0,85	0,99	0,94	0,99	0,98	0,69	0,89	0,9	0,85	0,99	0,91
0,76	0,99	0,61	0,79	0,98	0,85	0,99	0,94	0,98	0,99	0,91	0,96	0,95	0,89	0,98	0,94	0,93	0,71	0,93	0,86
0,83	0,99	0,81	0,88	0,86	0,84	0,83	0,84	0,98	0,83	0,84	0,88	0,73	0,94	0,99	0,89	0,95	0,99	0,98	0,97
0,92	0,89	0,99	0,93	0,89	0,73	0,86	0,83	0,94	0,83	0,86	0,88	0,99	0,72	0,93	0,88	0,93	0,96	0,99	0,96
0,94	0,96	0,68	0,86	0,99	0,93	0,99	0,97	0,91	0,99	0,95	0,95	0,99	0,99	0,88	0,95	0,9	0,99	0,79	0,89
0,97	0,96	0,91	0,95	0,89	0,95	0,99	0,94	0,95	0,99	0,99	0,98	0,99	0,91	0,86	0,92	0,98	0,56	0,86	0,80
0,5	0,7	0,96	0,72	0,97	0,88	0,98	0,94	0,67	0,52	0,61	0,60	0,76	0,98	0,99	0,91	0,96	0,87	0,8	0,88
0,86	0,98	0,99	0,94	0,99	0,96	0,93	0,96	0,92	0,69	0,99	0,87	0,74	0,64	0,88	0,75	0,93	0,87	0,92	0,91
0,86	0,99	0,98	0,94	0,99	0,93	0,69	0,87	0,71	0,98	0,93	0,87	0,98	0,79	0,99	0,92	0,89	0,69	0,86	0,81

Note. Each row of Table 4 represents a simulated case, composed of its respective values for the three components of the generalised K-expert coefficient: the objective dimension (K_c), the subjective dimension (K_a) and the contextual dimension (K_b) considering equal weighting.

It is important to note that Tables 2, 3 and 4 present simulated scenarios constructed for illustrative and methodological purposes only. These examples do not correspond to real empirical data, but to hypothetical configurations designed to examine the behaviour of the coefficient under different weighting schemes. In particular, these tables allow us to visualise the effect of classical co-weighting versus configurations with differentiated weights, showing how small variations in the weighting structure can modify the ranking of experts. The purpose of these scenarios is not to estimate population parameters, but to understand the structural impact of the equipondering assumption on the determination of the K coefficient.

Real-world application

Recent research has highlighted the need to adapt digital teacher competence frameworks to specific initial teacher education contexts, particularly in the field of pre-school education. In this framework, the study by Silva-Quiroz et al. (2025) proposed an adaptation of the European DigCompEdu framework to the context of students in initial teacher training for early childhood education in Chile, validating this adaptation through expert judgement. For this purpose, the expert competence coefficient (K) was used, selecting judges based on their self-assessment of knowledge (Kc) and argumentation of their expertise (Ka). However, such an approach, while useful, had limitations associated with the equal weighting of its components and the absence of robust inferential criteria to support the validity of the process.

Based on this real case, this paper proposes a reformulation in the use of the K coefficient, incorporating differentiated weights and a new component of thematic proximity, together with an inferential model based on the beta distribution and the asymptotic normal approximation. For the application of this proposal, 32 potential experts were involved in the validation process of the instrument. From the classical perspective, the results of the K coefficient estimation are presented in Table 5, which shows the classification of the experts according to the score obtained in Silva-Quiroz et al. (2025).

Table 5

Expert competence coefficient (K) obtained by each of the experts in Silva-Quiroz et al. (2025)

Person	Knowledge coefficient (Kc)	Argumentation coefficient (Ka)	Expert Competence Coefficient (K)
1	0,92	0,8	0,86
2	0,78	0,8	0,79
3	0,78	0,9	0,84
4	0,84	0,8	0,82

5	0,92	0,7	0,81
6	0,78	0,6	0,69
7	0,76	0,7	0,73
8	0,78	0,7	0,74
9	0,90	0,8	0,85
10	0,83	0,7	0,77
11	0,85	0,7	0,77
12	0,85	0,9	0,87
13	0,92	0,9	0,91
14	0,91	1	0,96
15	0,83	0,8	0,81
16	0,70	0,5	0,60
17	0,90	1	0,95
18	0,77	0,8	0,78
19	0,84	0,9	0,87
20	0,90	0,8	0,85
21	0,91	1	0,96
22	0,92	0,8	0,86
23	0,83	0,7	0,77
24	0,55	0,5	0,53
25	0,92	1	0,96
26	0,91	0,8	0,85
27	0,91	0,8	0,85
28	0,75	0,6	0,68
29	0,78	0,7	0,74
30	0,78	0,7	0,74
31	0,70	0,5	0,60
32	0,75	0,6	0,68

Note. Source: Silva-Quiroz et al. (2025).

Under the classical selection criteria, only those potential experts who obtained a coefficient $k \geq 0.8$, (consider approximations) will be participants in the validation process, characterising a sample of 22 experts.

Now, based on the proposal, the confidence interval for the sample of potential experts is (0.8310; 0.9462), establishing that all the values of the confidence interval are above the selection threshold, however, under the proposed method, the analysis sample does not constitute a valid sample of experts as 1 no pertenece al $IC\left(\frac{p}{p+q}\right)$, therefore a sub-sample should be selected from the potential experts, which characterises the group to participate in the validation process, to which it is established that there are only 15 experts who could participate in the validation process, as only there is obtained the interval (0.8696; 1.0000), which would characterise a sample of experts under the proposed criteria.

Finally, a test for paired samples is performed in order to determine whether the indifferenciation in the scores of the four descriptors of

K according to the proposal of Dobrov and Smirnov (1972), compared to the current proposal presented in Table 6, produces significant differences in the determination of the coefficient *K*. For this purpose, the following weightings are considered.

Table 6

Proposed weights for the criteria defining K_a

Indicator	Observation	High	Medium	Low
Indicator 1	Weighting identical to source	0.30	0.20	0.10
Indicator 2	Weighting identical to source	0.30	0.20	0.10
Indicator 3	Change in weighting	0.10	0.075	0.025
Indicator 4	Change in weighting	0.10	0.075	0.025
Indicator 5	Change in weighting	0.10	0.075	0.025
Indicator 6	Change in weighting	0.10	0.075	0.025

The results are presented in Table 7, after verifying compliance with the normality assumption using the Shapiro-Wilk test ($p = 0.261$).

Table 7

Paired samples t-test

Comparison	Statistic t	gl	p
K-Conventional vs. K-Proposal	51.1	31	< .001

Note. $H_a: \mu_1 - \mu_2 \neq 0$ (significant difference between measures).

From the preliminary contrast, it is concluded that there are significant differences, which is evidence of an overestimation attributable to undifferentiation in the weights. Considering the weights proposed in this research, it is observed that the mean of the *K* coefficient in its conventional version is 0.823, while in the proposed version it reaches a mean value of 0.754. See Table 8.

Table 8

Descriptive statistics of K-Conventional and K-Proposed

Measure	N	Mean	Median	SD	ES
K-Conventional	32	0.823	0.850	0.122	0.0216
K-Proposal	32	0.754	0.775	0.127	0.0224

Conclusion

The selection of experts in validation processes is a critical component in the construction of standardised measurement and evaluation instruments or in the abstraction of social phenomena, since the instruments to be used in the research are based on this choice. However, this procedure has been questioned, especially from methodological approaches that demand greater statistical rigor. In response, proposals have emerged such as the *k*-expert coefficient, which offers a quantitative alternative for identifying valid experts from a sample of potential candidates. This coefficient integrates two dimensions: the objective, linked to the verifiable fulfilment of certain descriptors, and the subjective, understood as the self-assessment of one's own competence.

However, the original model presents a relevant limitation by giving the same weight to both dimensions through a simple average, which is problematic considering the cognitive bias evidenced in the Dunning-Kruger effect (Mazor and Fleming, 2021), which shows that individuals with lower competence tend to overestimate their skills. This observation motivates the need to review the structure of the coefficient, especially with regard to the weighting of its components.

Given this scenario, a generalisation of the coefficient *k* is proposed that incorporates a weighted function with three components, under the restriction that the sum of their weights is equal to one. This formulation gives different weights to the objective dimension (verifiable criteria), the subjective dimension (self-assessment) and adds a new factor: the proximity of the expert to the context or target audience of the study.

This generalised model includes as a particular case the classic formulation (when $\alpha = 0.5$, $\beta = 0.5$ and $\gamma = 0$), but it is configured as a more flexible methodological framework that can be adapted to different research contexts.

Additionally, the coefficient *k* is redefined as a random variable with a beta distribution, which enables the use of statistical inference tools (such as parameter estimation and confidence intervals) to inform expert selection. This approach replaces reliance on arbitrary thresholds, such as the fixed value of 0.8, and allows decisions based on empirical evidence and statistical significance.

In parallel, the scoring scale for the five descriptors of the objective component is revised, as the original formulation showed limitations of discrimination and violated the principle of parsimony. The new scale aims to improve accuracy and consistency with the fundamentals of educational measurement.

The results obtained in real applications support the relevance of this methodological proposal. The implementation of the generalised *k* coefficient showed a significant improvement in the discrimination between potential experts, especially in situations where the self-assessment was inflated. The incorporation of the γ component, linked to the contextual affinity of the expert with the phenomenon under study, made it possible to adjust the selection towards more relevant profiles, without relying exclusively on formal or self-declared criteria. In addition, the use of statistical inference to estimate confidence intervals made the process more robust, replacing the fixed cut-off logic.

These results indicate that the proposed methodology is statistically sound, operationally more accurate and conceptually more aligned with current requirements in validation processes.

In summary, the generalisation of the k coefficient contributes to a more rigorous and transparent expert selection process, and opens up new lines of research aimed at optimising the weighting of its components according to the type of study, discipline or target population. This progress consolidates a robust and adaptable methodological framework, strengthening the validity of expert judgements and, by extension, the quality of scientific evaluation processes.

The results of the study confirm that the revised version of the k coefficient offers substantive advantages. In particular, the modelling of the subjective component by means of a beta distribution, together with the use of a normal approximation to define the threshold of "expertise", allows confidence intervals to be derived and hypothesis testing on the assessed competence. This inferential capability adds objectivity to the process. As Cabero-Almenara et al. (2020) have pointed out, the classical technique of expert competence already allowed for adequate discrimination of experts, and the proposed model strengthens this capacity with a more robust statistical basis. In line with this, Marín-González et al. (2021) found that the k coefficient effectively validated the analysed design, allowing for the identification of highly specialised experts.

The proposed methodological innovation is of a general nature and can be applied in other fields, such as higher education, health, engineering, among others, where expert judgement is required. The beta-normal formulation is applicable as long as the judgments are bounded in the interval [0,1]; therefore, future research could adapt the model to different scales and methodologies such as Delphi. Also, the possibility of assigning differentiated weights opens up new lines of work. Dobrov and Smirnov (1972) already pointed out that it is not inappropriate to assign different weights to Kc and Ka according to their nature. It would be interesting to explore Bayesian approaches or alternative consensus mechanisms to calibrate these weights, as well as to assess the stability of the coefficient in larger samples (Zartha Sossa et al., 2014). Finally, as Maldonado-Suárez and Santoyo-Telles (2024) emphasise, it is crucial to systematise and document the entire expert judgement process to ensure its rigour and traceability. Following these recommendations, through detailed reporting of the statistical and empirical criteria used, will contribute to strengthening the validity of future applications of the k coefficient in educational and scientific contexts. While the inferential approach introduces greater rigour with respect to the traditional fixed threshold, it should be recognised that the use of the confidence interval as a decision criterion may lead to overly categorical interpretations if not properly contextualised. Whether or not the reference value belongs to the interval is not a definitive validation, but a probabilistic approximation conditioned by model assumptions and sample variability. Therefore, the stochastic criterion should be understood as complementary and not a substitute for methodological judgement, avoiding rigid dichotomous interpretations in the

classification of experts.

Although the results obtained suggest that the proposed reformulation improves internal consistency and accuracy in the selection of experts with respect to the classical model, this evidence should be interpreted with caution. The scope of the conclusions is conditioned by the sample size, the distributional assumptions adopted and the illustrative nature of the simulated case. Consequently, rather than establishing a definitive superiority of the generalised model, the findings indicate its methodological potential as a more flexible and probabilistically grounded alternative. Future research with larger samples and diverse contexts will allow us to further assess its stability and discriminative power.

Funding

This research was funded by the Agencia Nacional de Investigación y Desarrollo (ANID) FONDECYT Regular N° 1230754.

References

- Aranda, T. M., Cantabrana, J. L. L., and Cervera, M. G. (2022). A technological solution for people with intellectual disabilities in emergency situations. *REICE: Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 20(2), 65-84. <https://doi.org/10.15366/reice2022.20.2.004>
- Barroso Osuna, J. M., and Cabero Almenara, J. (2013). The use of expert judgement for ICT evaluation: The expert competence coefficient. *Bordón. Journal of Pedagogy*, 65(2), 25-38. <https://doi.org/10.13042/Bordon.2013.65202>
- Cabero-Almenara, J., Barroso-Osuna, J., Rodríguez-García, A. M., & Llorente-Cejudo, C. (2020). Expert judgement in the validation of assessment instruments: The Delphi technique. *Education in the Knowledge Society (EKS)*, 21, e21923. <https://doi.org/10.14201/eks.21923>
- Capdevila-Pons, M., Pla-Campas, G., and Gironès, X. (2023). An early, intensive and complementary aquatic rehabilitation protocol after arthroscopy rotator cuff repair: Consensus through a Delphi study. *Advances in Rehabilitation / Postępy Rehabilitacji*, 37(1). <https://doi.org/10.5114/areh.2023.124852>
- Carrasco, P., and Dois, A. (2020). Competency profile of the nursing clinical tutor from the perspective of expert nursing staff. *FEM: Journal of the Medical Education Foundation*, 23(2), 81-87. <https://doi.org/10.33588/fem.232.1048>
- Cerero, J. F. F., Batanero, J. M. F., and Almenara, J. C. (2023). Digital teaching competencies and disability: Validation of a questionnaire design using the K coefficient to select experts. *Heliyon*, 9(6), e16467. <https://doi.org/10.1016/j.heliyon.2023.e16467>
- Dobrov, G. M., and Smirnov, L. P. (1972). Forecasting as a means for scientific and technological policy control. *Technological Forecasting and Social Change*, 4(1), 5-18. [https://doi.org/10.1016/0040-1625\(72\)90043-1](https://doi.org/10.1016/0040-1625(72)90043-1)

- Estrada Girón, A., Moreno Tapia, J., and Cáceres Mesa, M. L. (2022). Self-diagnosis of reading skills, metacognition and the Dunning-Kruger effect in high school students. *Revista Metropolitana De Ciencias Aplicadas*, 5(3), 105-112. <https://doi.org/10.62452/nmpydx77>
- Fernández Batanero, J. M., and Reyes Rebollo, M. M. (2016). Entrepreneurial skills of adult continuing education students in Andalusia. Teachers' perceptions. *Educación XX1*, 20(1). <https://doi.org/10.5944/educxx1.17511>
- Garrido-García, R., López-Esteban, P., and Martínez-González, R. (2020). Competencies and functions of the nursing clinical tutor: Analysis using Delphi technique. *FEM: Journal of the Medical Education Foundation*, 23(1), 12-19. <https://doi.org/10.35669/fem2020.23.1.12>
- Gelman, A., & Hennig, C. (2017). Beyond subjective and objective in statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4), 967-1033. <https://doi.org/10.1111/rssa.12276>
- González, J., Galvis, D., & Hurtado, L. (2014). The Generalized Beta distribution as a survival model to analyze university evasion. *Estudios pedagógicos (Valdivia)*, 40(1), 133-144. <http://doi.org/10.4067/S0718-07052014000100008>
- Herrera Masó, J. R., Calero-Ricardo, J. L., González Rangel, M. Ángel, Collazo Ramos, M. I., and Travieso González, Y. (2022). The expert consultation method at three levels of validation. *Revista Habanera De Ciencias Médicas*, 21(1), e4711. <https://revhabanera.sld.cu/index.php/rhab/article/view/4711>
- Juárez-Hernández, L. G., & Tobón, S. (2018). Analysis of the implicit elements in the content validation of a research instrument. *Revista Espacios*, 39(53). <https://www.revistaespacios.com/a18v39n53/a18v39n53p17.pdf>
- Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Maldonado-Suárez, N., & Santoyo-Telles, F. (2024). Content validity by expert judgement: Quantitative and qualitative integration in the construction of measurement instruments. *REIRE Revista d'Innovació I Recerca En Educació*, 17(2), 1-19. <https://doi.org/10.1344/reire.46238>
- Marín-González, F., Pérez-González, J., Senior-Naveda, A., and García-Guliany, J. (2021). Validation of the design of a scientific-technological cooperation network using the K coefficient for the selection of experts. *Información Tecnológica*, 32(2), 79-88. <https://doi.org/10.4067/S0718-07642021000200079>
- Mazor, M., and Fleming, S. M. (2021). The Dunning-Kruger effect revisited. *Nature Human Behaviour*, 5(6), 677-678. <https://doi.org/10.1038/s41562-021-01101-z>
- Ramírez, M. C., and Cepena, M. C. M. (2020). Origin and development of an index of expert competence: The K coefficient. *Revista Latinoamericana de Metodología de la Social Research*, (19), 40-56.

- Silva-Quiroz, J., González-Campos, J., Garrido-Miranda, J., Lázaro-Cantabrana, J., and Canales-Reyes, R. (2025). Adapting and validating DigCompEdu for early childhood education students through Expert Competence Coefficient. *Social Sciences*, 14(6), 345. <https://doi.org/10.3390/socsci14060345>
- Torres, J. A., & Fernández, J. M. (2015). Promoting inclusive schools: analysis of teachers' perceptions and needs from an organisational, curricular and professional development perspective. *Revista Electrónica Interuniversitaria Interuniversitaria de Formación del Profesorado*, 18(1), 177-200. <https://doi.org/10.6018/reifop.18.1.214391>
- Vílchez, J. L. (2020). Unnecessary dependence on the teacher: Results of systematic correction on the Dunning Kruger effect in an educational context. *Transformación*, 16(3), 453-464. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-29552020000300453&lng=es&tlng=es.
- Zartha Sossa, J. W., Montes Hincapié, J. M., Toro Jaramillo, I. D., & Villada, H. S. (2014). Delphi method: Proposal for the calculation of the number of experts in a Delphi study on biodegradable packaging to 2032. *Spaces Magazine*, 35(13), 10. <https://www.revistaespacios.com/a14v35n13/14351310.html>

Traducido con  DeepL

Date received: 12 September, 2025

Revision date: 22 September, 2025

Acceptance date: 3 March, 2026