REVISTA INVESTIGACION OPERACIONAL
https://doi.org/10.5281/zenodo.23302384 **VOL. 48, NO.1, 95-101, 2027**

GENERIC COMPETENCES AND HOLISTIC EDUCATION IN ACCOUNTING STUDENTS: EVALUATION OF A STRUCTURAL EQUATION MODEL

Domingo Guzmán Chumpitaz Ramos[*1], Víctor Ricardo Masuda Toyofuku[**], Catya Evelyn Vásquez Tarazona[**], Héctor Félix Cerna Maguiña[**], David Cirilo León Martínez[**]

[*]Universidad Nacional del Santa, Escuela de Posgrado, Nuevo Chimbote, Perú
[**]Universidad Nacional Mayor de San Marcos, Lima, Perú.

ABSTRACT

This paper evaluates a structural equation model linking five generic competences to the holistic education of final-year Accounting students at a Peruvian public university (N = 193). Estimates are taken from the underlying doctoral thesis; nothing is re-estimated here. What this paper adds is a full audit of what the reported figures do and do not support. Recomputing the free-parameter count from AIC = chi-square + 2q shows that the four rival specifications were fitted to 27, 25, 25, 21 observed variables respectively, and the final model to 17: the specifications do not share a data matrix, so their AIC values are not mutually comparable and the decreasing AIC does not establish superiority. Two of the five competences and one full indicator of each endogenous construct were removed along the way. In the final model, research and leadership predict the mediating construct positively, whereas risk and disaster management does so NEGATIVELY, a sign that the source reports and that is made explicit here. Holistic education, with an average variance extracted of 0.431, falls below the conventional convergent-validity threshold and is retained on two indicators only. The paper reports what the evidence sustains and lists what remains unverifiable.

KEYWORDS: structural equation modeling; model comparability; Akaike information criterion; generic competences; holistic education; measurement validity.

MSC: 62H25; 62P25

RESUMEN

Se evalúa un modelo de ecuaciones estructurales que relaciona cinco competencias genéricas con la formación integral de estudiantes del último año de Contabilidad de una universidad pública peruana (N = 193). Las estimaciones proceden de la tesis doctoral de base; aquí no se reestima nada. Lo que este artículo aporta es una auditoría de qué sostienen y qué no sostienen las cifras publicadas. Recomputando el número de parámetros libres a partir de AIC = χ^2 + 2q se obtiene que las cuatro especificaciones rivales se ajustaron sobre 27, 25, 25, 21 variables observadas respectivamente, y el modelo final sobre 17: las especificaciones no comparten matriz de datos, por lo que sus AIC no son comparables entre sí y el descenso del AIC no demuestra superioridad. Por el camino se retiraron dos de las cinco competencias y un indicador de cada constructo endógeno. En el modelo final, investigación y liderazgo predicen positivamente el constructo mediador, mientras que gestión de riesgos y desastres lo hace de forma NEGATIVA, signo que la fuente reporta y que aquí se explicita. La formación integral, con una varianza media extraída de 0,431, queda por debajo del umbral convencional de validez convergente y se retiene sobre solo dos indicadores. Se reporta lo que la evidencia sostiene y se enumera lo que no ha podido verificarse.

PALABRAS CLAVE: ecuaciones estructurales; comparabilidad de modelos; criterio de información de Akaike; competencias genéricas; formación integral; validez de medida.

1. INTRODUCTION

Holistic education —student development beyond the technical mastery of one's discipline— is a stated goal of Latin American higher education, but its measurement and the identification of its determinants remain open questions. This paper evaluates a structural equation model estimated on Accounting students at a Peruvian public university, derived from a doctoral study [4].

[1]Corresponding author: dchumpitazr@unmsm.edu.pe

Prior work in the region links dimensions of the formative process to student performance. Estupiñán Ricardo, Leyva Vázquez and Romero Fernández [6] examine, at an Ecuadorian university, the association between the motivation, organization and interaction-collaboration dimensions of project-based learning and academic performance, using Spearman correlation coefficients and Cronbach's alpha reliability. The present study shares with that one an interest in linking formative dimensions to student outcomes within the Latin American university context, but differs in its variables —generic competences and holistic education, not classroom methodology and performance—, in its unit of analysis and in its method: here a latent-variable model with a mediation structure is estimated, not bivariate correlations. That antecedent situates the problem within the regional literature; it does not validate the model evaluated here.

This paper's contribution is one of audit, not of methodological novelty. The analysis works on the estimates published in the doctoral source, which constitute its starting material; no specification is re-estimated. The procedure consists in recomputing, from those figures, the algebraic identities that a covariance model must satisfy, and using them to examine what conclusions the figures actually support. The main result of that audit is negative and is reported as such: the four rival specifications are not comparable by information criteria, because they were not fitted on the same set of observed variables.

2. METHOD AND DATA SOURCE

2.1 Design and sample

Quantitative, non-experimental, cross-sectional design. Population: 600 students of the Professional School of Accounting, Faculty of Accounting Sciences, Universidad Nacional Mayor de San Marcos, 2025 cohort. Instrument of 28 Likert-type items; Cronbach's α reliability of 0.9303 according to the source [4]. The source alternates between 2023 and 2025 when referring to the cohort; 2025 is adopted here because it is the year recorded in the operationalization matrix and in the statement of objectives and hypotheses, and it is the predominant one in the document (59 mentions against 11). The occurrences of 2023 are concentrated in the abstract and appear to be a residue of an earlier version.

On sample size, the source alternates between two figures, 185 and 193. The question can be settled arithmetically without access to the microdata. For a covariance model, the root mean square error of approximation satisfies the identity in equation (1) [17], from which the sample size can be solved for. Applied to the four rival specifications, it returns N = 193 for all four, and for the final model as well. The narrative according to which the sample was 185 and was expanded to 193 only for the estimation of the final model is therefore not compatible with the published indices: all reported specifications already correspond to N = 193.

$$\mathrm{RMSEA} = \sqrt{\frac{\chi^2 - \mathrm{gl}}{\mathrm{gl}(N-1)}} \tag{1}$$

It is not claimed here that the sample was expanded to improve fit: the source does not document that motivation. What is established is the discrepancy between the two figures, and the analysis is conducted on the one that the published indices imply.

2.2 Specification

In standard latent-variable notation [10, 11], the measurement model relates indicators to constructs and the structural model relates the constructs to one another. The final specification is a two-stage mediation: the exogenous competences predict a mediating construct, generic competences, and only this construct has a path to holistic education, equation (2).

$$\eta_1 = \Gamma\xi + \zeta_1, \ \eta_2 = \beta_{21}\eta_1 + \zeta_2 \tag{2}$$

The status of that mediation should be made precise. That no competence has a direct path to holistic education is a specification decision, not a finding: the model imposes mediation rather than testing it. Establishing mediation would require comparing this specification with an alternative that includes the direct paths, and testing the indirect effects

with appropriate intervals —distribution of the product or resampling [13]—, which requires microdata that are not available. The point indirect effects are reported in Section 3.4 without an interval, and full mediation is not claimed.

Estimation was by maximum likelihood in AMOS 25; normality was assessed with LISREL 8.72 [10]. The items are ordinal Likert-scale items, and the source reports a partial violation of normality. Consulting several fit indices does not make the estimation robust: an adequate treatment would require an estimator for ordinal variables or robust errors [11], which was not done and cannot be remedied without the data. It is declared here as a limitation of the source.

3. RESULTS AND AUDIT

3.1 Comparability of rival specifications

The source estimates four successive specifications and selects the one with the lowest AIC. That reasoning requires the models to be comparable. Table 1 shows that they are not, and it shows this with elementary arithmetic.

In AMOS output the Akaike criterion is computed as in equation (3), where q is the number of free parameters [1]. It should be noted that this is the particular form taken by certain SEM outputs; the general definition of the criterion is $-2 \log L + 2q$ [3], and the two coincide up to a constant that depends only on the data. Solving (3) for q. Furthermore, in a covariance model without a mean structure, the number of non-redundant elements of the observed matrix satisfies equation (4), where p is the number of observed variables.

$$\mathrm{AIC} = \chi^2 + 2q \tag{3}$$

$$\mathrm{gl} + q = \frac{p(p+1)}{2} \tag{4}$$

Specification	χ^2 (df)	AIC	q	df + q	observed p	N
Model 1	504.98 (308)	644.98	70	378	27	193
Model 2	430.32 (259)	562.32	66	325	25	193
Model 3	414.95 (257)	550.95	68	325	25	193
Model 4	308.31 (176)	418.31	55	231	21	193
Final model	203.73 (110)	289.73	43	153	17	193

Table 1. Free parameters, observed variables and sample size implied by the published figures. q is solved for from equation (3), p from equation (4) and N from equation (1).

The value of p comes out as an exact integer in all five specifications, which confirms identity (4) and, with it, that the number of observed variables changes across models: 27 → 25 → 25 → 21 → 17. The instrument declares 28 items and the source publishes loadings for 27. The trajectory is consistent with the successive elimination documented in the source itself: first two low-loading items, then the entire social-responsibility construct, and finally the ICT-management construct.

The consequence is direct. AIC can only rank models fitted to the same data matrix. Of the five specifications, only Model 2 and Model 3 share p = 25 and are therefore comparable to each other; no other pair is. The monotonic decrease of AIC across the sequence does not demonstrate the superiority of any specification: it largely reflects the fact that fewer and fewer variables are being modelled. The four specifications should be read as stages of a pruning process, not as rival alternatives subjected to a selection criterion.

For this reason, the selection among specifications cannot be presented as a multicriteria decision contribution. A table of fit indices does not by itself constitute a multicriteria method, and it does not constitute one at all when the alternatives are not comparable.

3.2 Measurement model

Table 2 reports, for each construct, the composite reliability and average variance extracted computed from the standardized loadings and error variances published in the source, following the criteria of Fornell and Larcker [8].

Construct	k	Items	Composite reliability	AVE	Minimum loading
Research	4	X1, X2, X3, X4	0.8468	0.5822	0.65
Leadership	4	X5, X6, X7, X8	0.8246	0.5403	0.72
Social responsibility	4	X9, X10, X11, X12	0.8269	0.5445	0.71
Risk management	4	X13, X14, X15, X16	0.8558	0.5977	0.74
ICT management	4	X17, X18, X19, X20	0.8697	0.6259	0.73
Holistic education	3	Y1, Y2, Y3	0.6744	0.4311	0.35

Construct	k	Items	Composite reliability	AVE	Minimum loading
Generic competences	4	Y5, Y6, Y7, Y8	0.8104	0.5215	0.56

Table 2. Composite reliability and average variance extracted by construct, computed from the standardized loadings in the source.

All constructs exceed the conventional 0.70 threshold for composite reliability. The average variance extracted, however, leaves one construct out: holistic education reaches 0.4311, below the 0.50 that convention requires to sustain convergent validity [8]. The origin is identifiable: its indicator Y1 has a standardized loading of 0.35, well below the rest. That item is dropped in the final model, which resolves the loading problem but leaves the construct with two indicators. A two-indicator factor is only identified within a larger model and its solution is fragile [11]; this is stated here rather than omitted.

Assessing discriminant validity requires the matrix of correlations among constructs, which the source does not publish. It is therefore not claimed that the constructs are discriminant: the property remains unestablished.

3.3 Fit of the final model

Incremental indices compare the model against an independence model [2, 18]. Table 3 reports the fit of the final model. χ^2 and degrees of freedom are reported because without them AIC is not auditable.

Index	Value	Usual reference
χ^2 (df)	203.7316 (110)	—
χ^2/df	1.8521	< 3 (indicative)
CFI	0.9412	≥ 0.90
TLI	0.9273	≥ 0.90
IFI	0.9422	≥ 0.90
RMSEA (90 % CI)	0.0666 (0.0522–0.0808)	≤ 0.08
SRMR	0.0376	≤ 0.08
RMR	0.0353	close to 0
GFI / AGFI	0.8954 / 0.8545	≥ 0.90 (indicative)
AIC	289.7316	only comparable at equal p
Free parameters q	43	—
Observed variables p	17	—

Table 3. Fit of the final model. χ^2, df, q and p were recomputed from identities (1), (3) and (4) and from the published χ^2/df ratio; the rest comes directly from the source.

The thresholds in the third column are the conventionally used ones. They should not be indiscriminately attributed to Hu and Bentler [9]: those authors propose combinations of criteria —for instance CFI ≥ 0.95 together with SRMR ≤ 0.08— and warn against their use as universal thresholds. The values of 0.90 for CFI and TLI and of 0.08 for RMSEA come from an earlier, laxer convention, and are declared as such here.

3.4 Structural relationships

Table 4 reports the coefficients of the final model with their standard errors, critical ratios and Wald confidence intervals.

The magnitudes in Table 4 are not comparable to one another: being unstandardized coefficients, each one is on the original scale of its indicator, so their relative size does not indicate which relationship is stronger. Comparing constructs against each other would require the standardized coefficients, which the source does not publish (see Limitations).

Relationship	Estimate	SE	C.R.	p	CI 95 %
Generic competences ← Research	0.2850	0.1221	2.3343	0.0196	[0.0457; 0.5243]
Generic competences ← Leadership	0.6278	0.1219	5.1503	< 0.001	[0.3889; 0.8667]
Generic competences ← Risk management	-0.2240	0.1026	-2.1819	0.0291	[-0.4251; -0.0229]
Holistic education ← Generic competences	0.6379	0.0961	6.6381	< 0.001	[0.4495; 0.8263]

Table 4. Unstandardized structural coefficients of the final model. The intervals are Wald intervals, computed as estimate ± 1.96 · SE. The source does not publish the standardized coefficients of the structural model, so they are not reported.

The central result requires a qualification. Three of the five competences maintain a statistically significant association with the mediating construct, but not in the same direction: research and leadership do so with a positive sign, whereas risk and disaster management shows a negative coefficient (-0.2240, p = 0.0291). Presenting all three as equivalent contributions, without mentioning the sign, would be misleading. A negative sign in this position admits several readings —overlap between constructs, statistical suppression through collinearity with the other competences, or a substantive phenomenon— and none can be settled without the microdata.

For the two remaining competences, the correct formulation is that no statistical evidence of association was found in this specification, not that they «do not predict». A p-value above 0.05 is not proof of the absence of an effect, and dropping paths on that criterion alone is a pruning decision that should be declared as such.

The point indirect effects on holistic education, obtained as the product of coefficients, are: research 0.1818; leadership 0.4005; risk management -0.1429. They are reported without a confidence interval: obtaining one would require the distribution-of-the-product method or resampling [13] on the microdata, which are not available. This should not be read as a test of mediation.

An alternative that does not require microdata is the delta method [16]: for a two-stage indirect effect ab, with a and b the unstandardized structural coefficients and SE_a, SE_b their standard errors already published in Table 4, the first-order standard error is $SE(ab) \approx \sqrt{(b^2 \cdot SE_a^2 + a^2 \cdot SE_b^2)}$. Applied to the three indirect effects with b = 0.6379 (SE_b = 0.0961): research, ab = 0.1818, SE = 0.0826, z = 2.20, p = 0.028, 95% CI [0.020; 0.344]; leadership, ab = 0.4005, SE = 0.0984, z = 4.07, p < 0.001, 95% CI [0.208; 0.593]; risk and disaster management, ab = -0.1429, SE = 0.0689, z = -2.07, p = 0.038, 95% CI [-0.278; -0.008]. All three intervals exclude zero.

The delta method is a first-order approximation, known to behave conservatively or anti-conservatively in medium-sized samples relative to resampling or the distribution of the product [13], and it assumes independence between the errors of a and b, which cannot be verified without the covariance matrix of the estimators. It does not replace the resampling interval that Section 3.4 already identifies as the appropriate method; what it contributes is a bound calculable solely from the figures the source publishes, and, like the point effects it starts from, it should not be read as a test of mediation.

4. DISCUSSION

Substantively, the model suggests that not all of the declared competences feed equally into the mediating construct that leads to holistic education, and that one of them does so in the direction opposite to what was expected. For curricular design, the implication is not to abandon the competences with no detected association, but to review how they are taught and — above all — how they are measured: two of the seven constructs have indicators with standardized loadings below 0.60, and the dependent construct does not reach the conventional convergent-validity threshold.

Methodologically, the case leaves a lesson transferable to structural modelling practice, and that is the contribution this paper can actually sustain. When a sequence of specifications is pruned by removing indicators, information criteria cease to be comparable across stages, even though the software keeps printing them in the same column. The check is elementary — it suffices to solve for q from equation (3) and for p from equation (4) — and it should be part of the routine review of any paper that compares models by AIC.

An extension remains open that this work cannot carry out but that should be stated precisely. The relationships the model estimates are net associations between constructs; different from these are the relationships that faculty attribute among those same dimensions. Contrasting the two — the relationship map elicited from a faculty panel against the estimated associations — is an additional and legitimate question, and instruments exist to address it: the consensus process on mental models through fuzzy cognitive maps and computing with words of Pérez-Teruel, Leyva Vázquez and Estrada-Sentí [14] allows those attributions to be represented and the zones of disagreement to be located. It

requires collecting expert judgments, which do not exist here. It should be stressed that the structural coefficients are not the causal weights of a cognitive map and should not be transferred to one: they are distinct objects, and the comparison would be meaningful precisely because they can disagree. Likewise, a configurational analysis [15] on the same cases would answer a different, asymmetric question — which combinations of competences suffice — but it requires individual microdata and a justified calibration. Neither extension is undertaken here, and methods are not stacked up to feign novelty. Likewise, consideration of biases arising from institutional prestige or perception in the evaluation of student competences benefits from the analysis of institutional biases in evaluative models by Leyva Vázquez and Smarandache [12]. This perspective is complemented by the modelling of uncertainty in health sciences and higher education through neutrosophic logic [7], as well as by the evaluation of interdisciplinary competences under multivalued hypothesis theory [5].

5. LIMITATIONS

First: discriminant validity remains unestablished, since the source does not publish the matrix of correlations among constructs, and the dependent construct does not reach the conventional convergent-validity threshold. This limitation is given priority because it directly bounds what can be claimed about the distinction between the model's constructs, the intended use of its results.

Second and determinant: the starting material is the published estimates, not the microdata. Verification is therefore algebraic, and its scope is that of the internal consistency of the reported figures: it establishes which conclusions they support and which they do not, not whether the model is well specified.

Third: the design is cross-sectional and the reported relationships are associative, not causal, despite the source's use of the verb «influence».

Fourth: mediation is imposed by the specification and is not tested against a model with direct paths; the indirect effects are reported without an interval.

Fifth: maximum-likelihood estimation on ordinal indicators with violated normality was not corrected with an appropriate estimator.

Sixth: the exploratory and confirmatory analyses were carried out on the same sample, which gives the respecifications an exploratory character and limits the confirmatory interpretation of the final fit.

Seventh: the source does not publish the standardized coefficients of the structural model or the diagram of the final model, so both are omitted rather than reconstructed.

6. CONCLUSIONS

A structural equation model of generic competences and holistic education was audited from the figures published in its doctoral source, without access to microdata.

The main finding concerns comparability: the five reported specifications were fitted on 27, 25, 25, 21, 17 observed variables respectively, so their information criteria are not comparable with one another and the decrease in AIC does not establish the superiority of any of them. The specifications are stages of a pruning process, not rival alternatives.

The sample size implied by the published indices is N = 193 across all specifications, which does not agree with the narrative of a sample expansion carried out only for the final model.

In the final model, research and leadership are positively associated with the mediating construct and risk and disaster management is negatively associated with it; for the other two competences no evidence of association was found in this specification, which is not equivalent to the absence of an effect.

The dependent construct does not reach the conventional convergent-validity threshold and is left with two indicators in the final model. Discriminant validity could not be assessed.

RECEIVED: MAY, 2026.

REVISED: OCTOBER, 2026.

REFERENCES

[1]. AKAIKE, H. (1974): A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, 19(6), 716-723.

[2]. BENTLER, P.M. (1990): Comparative fit indexes in structural models. **Psychological Bulletin**, 107(2), 238-246.

[3]. BURNHAM, K.P. and ANDERSON, D.R. (2002): **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed.** Springer, New York.

[4]. CHUMPITAZ RAMOS, D.G. (2026): Modelo de ecuaciones estructurales para la formación integral de estudiantes de Contabilidad. Tesis doctoral, Universidad Nacional del Santa, Nuevo Chimbote, Perú.

[5]. DE LA CANTERA, D. H., QUIROZ, R. C., QUEIJA, M. L., GONZALEZ, J. R. and LEYVA VÁZQUEZ, M. Y. (2024): Analyzing Interdisciplinary Education in General Medicine Using Smarandache's Multivalued Logic Hypothesis Theory and Plithogenic Probability. **Neutrosophic Sets and Systems**, 70, 312-321.

[6]. ESTUPIÑÁN RICARDO, J., LEYVA VÁZQUEZ, M. Y. and ROMERO FERNÁNDEZ, A. (2022): Evaluación del aprendizaje basado en proyectos. **Revista Investigación Operacional**, 43(3), 409-419.

[7]. ESTUPIÑÁN RICARDO, J., LEYVA VÁZQUEZ, M. Y., ÁLVAREZ GÓMEZ, S. D., VELÁZQUEZ-SOTO, O. E. and RODRÍGUEZ-GUZMÁN, A. A. (2023): La aplicación de la neutrosofía en las ciencias médicas: una revisión bibliográfica narrativa. **Revista Cubana de Información en Ciencias de la Salud**, 34, e2599.

[8]. FORNELL, C. and LARCKER, D.F. (1981): Evaluating structural equation models with unobservable variables and measurement error. **Journal of Marketing Research**, 18(1), 39-50.

[9]. HU, L. and BENTLER, P.M. (1999): Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. **Structural Equation Modeling**, 6(1), 1-55.

[10]. JÖRESKOG, K.G. and SÖRBOM, D. (1996): **LISREL 8: User's Reference Guide**. Scientific Software International, Chicago.

[11]. KLINE, R.B. (2016): **Principles and Practice of Structural Equation Modeling, 4th ed.** Guilford Press, New York.

[12]. LEYVA VÁZQUEZ, M. Y. and SMARANDACHE, F. (2026): Institutional prestige bias in LLMs. **Neutrosophic Computing and Machine Learning**, 44, 54–63.

[13]. MACKINNON, D.P., LOCKWOOD, C.M. and WILLIAMS, J. (2004): Confidence limits for the indirect effect: distribution of the product and resampling methods. **Multivariate Behavioral Research**, 39(1), 99-128.

[14]. PÉREZ-TERUEL, K., LEYVA VÁZQUEZ, M. Y. and ESTRADA-SENTÍ, V. (2015): Mental models consensus process using fuzzy cognitive maps and computing with words. **Ingeniería y Universidad**, 19(1), 173-188. DOI 10.11144/Javeriana.iyu19-1.mmcp.

[15]. RAGIN, C.C. (2008): **Redesigning Social Inquiry: Fuzzy Sets and Beyond**. University of Chicago Press, Chicago.

[16]. SOBEL, M.E. (1982): Asymptotic confidence intervals for indirect effects in structural equation models. **Sociological Methodology**, 13, 290-312.

[17]. STEIGER, J.H. (1990): Structural model evaluation and modification: an interval estimation approach. **Multivariate Behavioral Research**, 25(2), 173-180.

[18]. TUCKER, L.R. and LEWIS, C. (1973): A reliability coefficient for maximum likelihood factor analysis. **Psychometrika**, 38(1), 1-10.