

**WHOSE PERCEPTION IS IT? RESPONDENT-LEVEL AND INSTITUTIONAL-
LEVEL EVIDENCE ON INTERNAL CONTROL QUALITY IN
MICROFINANCE INSTITUTIONS**

AMAH DENIS AGEM

PhD Candidate Accounting, Auditing, and Management Control at the Faculty of Economics and Applied Management of the University of Douala, Cameroon.
amahdenis2000@gmail.com

DR. CHI COLLINS PENN

Senior Lecturer, Department of Finance and Accountancy, Faculty of Economics and Applied Management, University of Douala, Cameroon
chingafor2000@yahoo.fr

Correspondence: Amah Denis Agem, PhD candidate, Accounting Audit and Management Control, Faculty of Economics and Applied Management, University of Douala, Cameroon.
amahdenis2000@gmail.com

DOI: <https://doi.org/10.63452/IJAFSSR.2026.4505>

ABSTRACT

Survey research on internal control asks individuals to describe a property of the organisation they work in and treats the answers as measurements of that organisation. This paper asks how much of such an answer reflects the organisation, and whether the two vantage points within an organisation describe the same thing. We surveyed 320 key informants in 80 Category 1 microfinance institutions in Cameroon, two from management and two from operational positions in each, with an institution code on every questionnaire. Between 4 and 13 percent of the variance in perceived control quality lies between institutions. The reliability of a four-informant institution mean is 0.39, and a single informant correlates at 0.59 with that mean. Management and operational respondents rate control equally favourably, and the association between internal control and information quality is almost identical in the two groups, yet their institution means correlate at only 0.34. Level and ordering come apart: a study surveying one position type would reach the same conclusion about the sector but a different one about which institutions are well controlled. We interpret these findings through composition model theory and argue that intraclass correlations should be reported as standard.

KEYWORDS: - Some Key Informants; Multi-Level Measurement Problems; On Composition Models; The ICC Revisited; Internal Controls and MFIs.

1.0 INTRODUCTION

There is very little that researchers wishing to learn about the quality of internal controls in organizations can actually do. Controls cannot be observed from the outside; their documentation concerns what was intended but does not reflect what occurs, and external ratings are available only for sufficiently important entities. Hence, they ask somebody working there.

That score becomes itself one measure of the organization. There is consequently much research showing relationships between internal controls measured thus and good reporting practices elsewhere in many places; see Phornlaphatrachakorn (2019); Monteiro et al. (2021); Hoai and Nguyen (2022); Tetteh et al. (2022); Nguyen et al. (2023). Such activity seems quite normal everywhere and certainly cannot be avoided anywhere. Few people seem curious about what proportion of each result derives from characteristics of organizations versus individuals responding.

There were actually two questions raised here. One was the variance question: of the differences among people concerning their own institution's control system, what proportion concerns differences among institutions themselves? Another regarded vantage point: would governors' descriptions agree with operators', if both described something they had together run? The latter could never emerge under any survey involving just one sort of roleholder; however, it proved surprisingly rewarding when answered.

There is no need for anything except that researchers record what organisation their respondents belong to and interview more than one representative. They performed both steps here by speaking separately to two managers and two operators at each of eighty institutions. Hence there are two sorts of contributions: first, the intraclass correlations themselves appear where before they did not; hence also some indication of the amount of organisational substance contained in any one person's own report taken alone; and secondly, comparison across positions inside each body itself reveals that although people can agree on where their degree of control stands overall, they disagree over their rankings relative to others'.

It consists of six sections. Section 2 explains the measurement theory underlying both questions and surveys the evidence. Section 3 describes the field site, experimental design, measures taken, and analytical approach adopted. Section 4 reports the findings. Section 5 discusses what was found and what was not, and relates these points to previous studies. Finally, some concluding remarks appear in section 6 along with comments about their meaning.

2.0 THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1 Composition Models and What They Require

It makes many distinctions about how a lower-level measure may represent some higher-level construct at all; these differences actually matter because they require different kinds of evidence.

With direct consensus modeling, there is no problem because under that kind of model, the higher-level construct consists precisely of what people have in common; hence, the unit's own measure depends entirely upon member agreement: if members agree, then what they say collectively makes sense. Referent-shift consensus modeling works similarly, except that questions refer to units instead of persons; many existing measures of internal controls appear to work this way. Additive models also present no difficulties, except here the unit value equals the mean alone; hence, there need never be any agreement at all, but neither does anything held mutually apply, since collectively one has only averages over individuals. Finally, dispersion approaches treat intraclass scatter itself as the object of study. Table 1 sets out the four alternatives, the evidence each requires, and what the present data provide.

Table 1: Composition Models and What Each Requires

Composition model	What the construct is	Aggregation justified by	Our evidence
Direct consensus	Something held by everyone who makes up the unit, on account of agreement among members.	Within-unit agreement (rwg, ICC)	Weak: ICC(1) = 0.136, ICC(2) = 0.39
Referent-shift consensus	One type of shared-characteristic question has elements referring to the unit rather than the respondent.	Within-unit agreement	Instrument uses unit referents; agreement still weak
Additive	The unit mean, irrespective of agreement	Nothing — agreement is not required	Defensible, but changes what is claimed
Dispersion	The variability within the unit is itself the construct	Theory about why variance matters	Not used here; a direction for further work

Note that classification follows the usual typology of compositional-model types found in multiple-level sorts of investigations generally; hence, in practice, there cannot be enough agreement for any kind of simple consensus if one takes the latter approach.

Survey research on internal control rarely states which model it adopts. Its language, describing the measure as the institution's control quality, implies consensus; its procedure, which aggregates or uses single informants without reporting agreement, is additive. The gap between the two is where this paper sits.

2.2 Two Reliabilities, and Why Both Are Needed

There are two indices. ICC(1) refers to the proportion of total variation residing among units; thus, its value indicates what portion of someone's response may be attributed to their unit. ICC(2) concerns the reliability of the unit mean; therefore, it also depends partly upon ICC(1). But these need not agree. With enough people living together, even small ICC(1)s produce fairly good unit means. Large ICC(1)s still yield poor unit-mean reliabilities unless there happen to be more than two persons per unit.

They are both required because each addresses a separate kind of design question: whether there will be one informant from each organization (ICC(1)) or more than one (ICC(2)). Neither value appears often here either, which means that nobody knows how precise any of these findings actually are.

2.3 Agreement and Its Ambiguity

Very low agreement among informants allows for either of these interpretations and rules out neither. One must therefore keep both possibilities alive until more evidence accumulates. Either members are describing the same system but doing so inconsistently, hence their lack of mutual accord reflects poor measurement, or else they truly have something else altogether under observation because what controls at a distant satellite has little relation to how control works where the centre itself makes decisions.

That is why it comes up once more in Section 5.1. This matters for interpretation; yet in what little previous work there was, consideration seems never to have been given to anything other than the former sense.

2.4 The Key Informant Tradition

In key-informant methods, individuals are selected according to their knowledge concerning what is being investigated instead of being representative of all workers, and there seems little reason to doubt this thinking: a general manager knows much more regarding controls themselves than does someone picked at random from among other employees. Most discussions of problems connected with taking measures via just one type of informant deal with common-methods bias (Podsakoff et al., 2003, 2024; Liang et al., 2007; Chin et al., 2012; Kock, 2015; Aguirre-Urreta & Hu, 2019). Common-methods variance refers precisely to relations between

two variables when both have been rated by one individual, while the issue raised here concerns whether anyone from anywhere inside an organization can give accurate information about some aspect of how things actually work throughout that entire place. One last point deserves attention: avoiding method bias will not solve these difficulties if they exist.

There has been much more attention paid generally by researchers studying organisations to multiple informants for many years in management and accounting studies than there appears to have been among internal control researchers in particular; indeed, the latter seem never actually even to have reported their own levels of agreement when they did use several respondents.

2.5 Empirical Review

There is also considerable evidence from actual practice concerning the connection between internal control and the quality of accounting information. Phornlaphatrachakorn (2019) finds that internal control quality affects the usefulness of accounting information, along with compliance and decision-making outcomes, among Thai food-processing companies. According to Monteiro et al. (2021), internal control and accounting systems influence the usefulness of financial information through their impact on its quality in Portugal; according to Nguyen et al. (2023), there is a correlation between internal control and the quality of accounts in quoted Vietnamese firms; according to Hoai and Nguyen (2022), internal control systems affect company profitability in developing countries, but only if leaders remain constant over time and maintain good quality themselves; finally, in line with the earlier findings on listed Ghanaian businesses by Tetteh et al. (2022), a similar pattern holds there, except that IT acts as a moderator instead.

There are two important features here concerning what was just described. One concerns the senses involved and their number. Most of these studies' measures concern perceptual matters, mostly based on information provided by one sense alone. Another has to do with units. Units claimed to have been studied (organizations themselves) differ from units actually observed (people working there). Chalmers et al. (2019) later reviewed research on IC but did not mention levels of measurement when explaining why certain problems arose if one wished to compare findings reported by various investigators.

There is less evidence concerning microfinance itself, and what there is concerns mostly governance rather than internal controls. In fact, Mersland and Strøm (2009) associate governance mechanisms with performance across microfinance institutions worldwide; Hermes and Hudon (2018) survey determinants of success throughout the entire field; Tadele et al. (2022) look at risk-taking and governance among microfinance providers in Sub-Saharan Africa; and Omidiji et al. (2024) describe internal auditing arrangements found in transition and

developing countries generally before concluding that great diversity prevails. Accountability seems problematic here as well, according to two reports about fulfilment of accountability requirements in ways that please outsiders but fail to enlighten their ostensible readers (Akanga, 2017; Marini et al., 2018). But again, nobody mentions an intraclass correlation coefficient, nor does anyone question more than one person per organization.

3.0 RESEARCH METHODOLOGY

3.1 Research Design and Philosophical Orientation

It is a cross-sectional survey carried out from a positivist standpoint. There is nothing interpretive about its questions. They concern proportions of variance attributable to organizations found on perception scales and degrees of agreement between kinds of informant inside organizations. Such things can be determined from distributions themselves once these have been measured and estimated. Hence they have nothing at all to do with causation, nor does anything else contained herein. Neither is there scope for causal inference owing to the nature of the data involved.

3.2 Study Area

It took place among Category 1 microfinance institutions in Cameroon. These are savings and credit cooperatives owned collectively by their members. Under rules made locally by the bank supervisory authorities themselves, these organizations have no choice but to take deposits from their own members, make loans solely to members, and join together in networks providing certain uniform practices along with some degree of monitoring; hence there is good reason why controls worthy of attention exist at all. In addition, they qualify because they hold elections for their governing boards among the membership generally (Cornforth, 2004). Therefore, people can describe what they know about how controls actually work while occupying quite distinct places relative to them. Lastly, they suffice because they will probably know everything important concerning any controls actually existing within their organizations.

3.3 Population, Sampling and Data Collection

It consists of Category 1 MFIs associated with one of these three networks (CAMCCUL, RECCUCAM, and RAINBOW-CAM) and carrying out activities in five Cameroonian regions (Centre, Littoral, North West, South West, and West). Altogether, eighty units selected at random according to stratified sampling constitute its sample. Networks serve here again as the basis for forming strata, which makes it possible almost to reproduce their relative proportions found when drawing up the list containing all members belonging to each network.

There are four key informants in each chosen institution. Two hold managerial posts (general manager, branch manager, internal controller, member of the board or supervisory board); two

others occupy operative positions (accountant, loan officer, cashier). This even distribution was built into the study design. Hence another sample emerges, consisting of some 320 individuals residing in roughly eighty organizations, with four in each organization studied.

There is also an institute code on every questionnaire supplied by the fieldworker prior to administration. A separate record linked these codes to institute names, but this link was destroyed afterwards, leaving nobody able to identify which institution any responses came from on the basis of anything reported here. Without this field nothing could have been computed, which is why none of the prior work discussed above reports equivalents of the statistics presented later.

That is why four informants were chosen from each of eighty institutions rather than more informants selected in fewer places. Whether one takes many cases with few observations per institution or few cases with more observations per institution influences how well between-unit variance will be estimated relative to mean values within units themselves. An optimization trade-off was made here. Its price appears later (in Sec. 6.3).

3.4 Variable Measurement

There were four activity dimensions based on COSO's own description of internal controls contained therein (COSO, Committee of Sponsoring Organizations of the Treadway Commission, 2013). These consisted of separation of duty, authorization, verification, and approval, review of work done, and reconciliation and auditing. There are two aspects concerning accounting information itself, namely relevance and reliability, modified according to van Beest et al. (2009); they consist entirely of statements concerning the university itself rather than anything relating directly to the individual who answered the questionnaire item in question, thus constituting Referent Shift types, explained above under Questionnaire Design Considerations. Internal control effectiveness itself is not a latent variable but rather a composite made up of its four activity components, because there is no such thing as good internal control unless something actually happens.

3.5 Validity and Reliability

It is deemed to have satisfactory content validity because its items come partly from the COSO framework itself and partly from van Beest et al.'s (2009) own measure of FRQ; in addition, some cognitive interviewing took place while developing the questionnaire, in order to check whether people understood each item as planned (Willis, 2005). Internal consistency and convergent validity pass their usual tests (Nunnally and Bernstein, 1994; Hair et al., 2019); discriminant validity has been checked by means of the HTMT criterion suggested recently by Henseler et al. (2015). Loadings vary between .69 and .89.

3.6 Analytical Strategy

There is also variance decomposition. It applies to construct scores yielded by PLS path modeling (Ringle et al., 2024); therefore, it is worth paying attention to, because there are three reasons why this matters.

One is that such a partition depends heavily upon what measurement noise the measure carries. This matters because taking an equally weighted average over all elements assumes they each convey equal information. With uneven qualities among them (loadings run from 0.69 to 0.89), unnecessary scatter arises unless one gives greater consideration according to each element's strength when forming its component totals. Only then will there be enough reduction of unwanted disturbance within units before comparing measures across them.

There is then secondly its nature: internal control effectiveness cannot reasonably be treated as some kind of hidden entity behind what an organization actually does concerning these matters but is instead made up by them. Hence, again, there is no room here for reflection, only composition; accordingly, partial least squares can handle this without further ado, whereas covariance-based methods need extra identification conditions before they will work at all if one wants to stick with this conception of the construct itself (Sarstedt et al., 2019, 2022; Hair et al., 2022).

3 Comparability. The literature addressed here fits its models via PLS, and carrying out the desired kind of decomposition based on scores produced by some other sort of estimator would mean obtaining results concerning something other than what these authors themselves have found. One's aim instead ought once again simply to describe the quantities they themselves have measured.

One-way random-effects ICC inferences depend on bootstrapping institutions instead of individuals for this purpose (as explained in Streukens and Leroi-Werelds, 2016); there are 5,000 repetitions altogether because otherwise one could not honor the intra-class correlations reported here when calculating standard errors.

3.7 Robustness Checks

There are two ways of checking whether measurement choice drives results. First, intraclass correlations calculated on unweighted scale means are only somewhat smaller than those found above. Hence they provide another basis on which one could conclude that there is hardly any difference among institutions. Second, consistent partial least squares (Dijkstra and Henseler, 2015), a method designed to correct for attenuation, does not change the picture appreciably when applied here.

4.0 RESULTS AND DISCUSSION

4.1 Variance Decomposition of Perceived Control and Information Quality

Table 2 breaks down the variance of every construct into its two parts: between institutions and within institutions.

Table 2: Variance Decomposition of Perceived Control and Information Quality

Construct	Between institutions	Within institutions	ICC(2)	SD between	SD within
Segregation of duties	12.6%	87.4%	0.37	0.588	0.869
Authorization, verification and approval	12.6%	87.4%	0.37	0.588	0.829
Review of task performance	11.3%	88.7%	0.34	0.580	0.866
Reconciliation and audit	4.4%	95.6%	0.16	0.533	0.896
Relevance	8.0%	92.0%	0.26	0.558	0.861
Reliability	6.3%	93.7%	0.21	0.546	0.880
Internal control (composite)	13.6%	86.4%	0.39	0.595	0.834
Accounting information quality	8.6%	91.4%	0.27	0.562	0.872

Note: N=320 individuals at 80 institutions; ICC(2) refers to reliability when using means based on four people taken from any one institution. Because scores have been standardized, however, SDs refer to their magnitude relative to individual differences overall.

Source: Author's computation, 2026

Between 4 and 13 percent of the variance in the individual dimensions lies between institutions, and 86 percent of the variance in the internal control composite lies within them. The reliability of a four-informant institution mean is 0.39 for that composite, well below the 0.70 conventionally sought before a group mean is treated as a measure of a group property. Reconciliation and audit shows the weakest between-institution component at 4.4 per cent, and internal control as a whole the strongest at 13.6 percent.

4.2 What a Single Informant Represents

This is the translation of the decomposition into quantities relevant for research design.

Table 3: What a Single Informant Represents

Quantity	Value	Interpretation
Mean difference between two respondents in the same institution	1.02 SD	How far apart two colleagues describing one control system are
Mean difference between two	1.13 SD	The same quantity with no shared institution

respondents in different institutions		
Ratio of the two	0.91	Sharing an institution reduces disagreement by about a tenth
The correlation between one randomly picked respondent and the institution mean based on four.	0.59	What a single-informant design captures of the institution
Relationship between the management mean of an institution and its own operational mean	0.34	Agreement between the two vantage points within an institution
Mean difference, management minus operational	-0.034 SD	Neither group rates control systematically higher ($p = .717$)

Note: All quantities computed on the standardized internal control composite.

Source: Author's computation, 2026

The difference between descriptions produced by two colleagues within one organization is 1.02 SD, compared with 1.13 SD when they come from separate institutions. Sharing an institution thus lowers their divergence by roughly ten percent. One randomly selected individual agrees with the average opinion of four others from the same institution to the extent of $r = +.59$.

Here they cannot mean that institutions make up most of the score obtained by any particular rater but only that their influence amounts to less than half when averaged over all raters. Thus understood, these numbers hardly rule out validity altogether, nor need one claim anything so extreme in order to advance the present point.

4.3 The Two Vantage Points: Level and Ordering

It allows making some comparisons impossible in studies based on reports by one informant alone. This is why it appears here as the second major result rather than because of considerations about variance partitioning.

They hardly diverge at all on either score. Management means vs. operations means for each institution: mean difference of 0.034 SD units, $p = .717$; relation of internal control to accounting information quality reported by management and supervisors compared with operators: 0.559 versus 0.566 respectively, with a difference of 0.007 (bootstrap CI includes zero comfortably). There appears to be no systematic tendency here toward neither exaggerated claims about good controls, nor any difference concerning what one can expect about the quality of accounts based on the amount of control present. It is therefore welcome news if one had feared managerial overstating of controls under their own jurisdiction.

There seems to be even less agreement when one considers order: the correlation coefficient for the correlation calculated from the means obtained by each manager compared with every operator at each institution is only 0.34. Thus there is hardly any relationship at all between these measures. Knowing that an institution's managers give their own control system high ratings does not help predict what operators will say about it.

Taken together, they tell much more than either says by itself. Had one investigator polled only managers and another only operator from the same organizations, they could have found perfect agreement about what degree of control prevailed overall in that kind of institution, and also about how strong its correlation with good information quality was. Yet they might still have disagreed greatly about which particular places were best governed. Levels agree but orders differ. Therefore, there is good reason here for caution: research based upon these indicators may safely describe a whole field, but if it seeks instead to rank-order individual cases found inside it, doubt arises. However, at present no distinction exists in the literature between these two situations.

One qualification is important here: the correlation is attenuated by the unreliability of a two-informer-group mean. Therefore, its true value must indeed exceed .34; precisely how much it exceeds that amount cannot be determined on the basis of two informers per group. Thus interpreted the finding shows that there is considerable difference between these orders but does not specify exactly what "considerable" means.

4.4 Consequences for Inference

Four respondents at once come from one institution, and four persons working in the same institution cannot provide four separate pieces of information. They observe the same control system and give rather similar answers; that is precisely measured by the intraclass correlation. Hence, the second, third, and even the fourth informant coming out of some particular unit contribute somewhat less novelty than does firstly appearing someone from somewhere else altogether. Ordinary formulas then calculate too much when they consider their input to consist of totally unconnected data points amounting in total number to 320.

That is what the design effect measures. With an intraclass correlation of 0.136 and four informants per institution, its value is $1 + (4 - 1) \times 0.136 = 1.41$, so variances are 41 percent greater than they would be under full independence. Dividing the sample size by the design effect yields an "effective" sample of 227: the 320 respondents contain roughly as much information as if 227 had answered independently, losing the equivalent of approximately 93 questionnaires because of repetition among cases within each unit.

They thus amount to approximately 19 percent more than if there had been no intra-institution correlation among the 320 observations. Hence confidence intervals grow wider and test statistics become smaller by factors proportional to this figure. Consequently, something that looked like a significant finding at the five-percent level becomes barely significant when accounted for adequately; this also explains the way the results presented above were inferred (see Section 3.6).

There is no punishment proportionate to the number actually interviewed, although that is what many people think at first. Had those same 320 questioned been allocated instead among either forty institutions (of eight each) or sixteen (twenty per institution), then the figures become respectively: design effect = 1.95, hence effective sample size = 164; design effect = 3.58, hence effective sample size = 89. Inter-interview contact thus helps if extra cases occur amongst additional institutions; otherwise there will still be some benefit from further interviews within the eighty places originally selected, though this declines greatly once started, since the largest possible effective sample equals the number of institutions divided by their intraclass correlation coefficient (about five hundred ninety in this case). In other words, contacting many individuals residing within very few organizations serves less well than contacting equally many spread over numerous establishments; hence, once again, the solution lies mainly in obtaining more units, instead of more persons.

There is no penalty according to how many informants there are altogether, although this last aspect seems hardest to grasp. Thus an intraclass correlation remains just what it was if based on fourteen individuals from each of twelve separate entities instead of seven apiece out of twenty-four. Taking several people per organization does nothing either when one has surveyed very many persons distributed among very few units in general.

5.0 DISCUSSION

5.1 What These Results Do and Do Not Show

It appears mostly to concern respondents rather than anything else on this occasion. One need some caution here, since there may appear to be more in this finding than actually exists.

It does not show that single-informant measures are invalid. A measure with a modest intraclass correlation can still correlate with outcomes, and the associations reported in the literature reviewed in Section 2.5 are not called into question by anything here. What it shows is that such a measure carries less information about the organisation than its usual description implies, and that the implied composition model, a shared property inferred from member agreement, is not supported by the agreement evidence.

Equally noteworthy, the result of the decomposition itself does nothing to show that people report what they know inaccurately. Another possibility was explained above (see section 2.3): their reports may reflect real differences because they actually learn about separate elements of the whole system while occupying their respective locations; hence they disagree, but for good reason. In fact, if there had been no connection whatsoever, one would have expected agreement between positions by chance alone, whereas the observed correlation coefficient of .34 suggests that some relationship was found despite weakening due to measurement error. Accordingly, more cannot be learned concerning how much of the discrepancy consists of mistakes versus true intra-institutional divergence from these data. Only if someone repeats the survey according to branches can progress occur here.

5.2 Comparison with Prior Literature

It remains true, according to what was just said about how internal control relates to accounting information quality, that this does not contradict anything found above but merely repeats it for yet another case none of these investigations considered (Phornlaphatrachakorn, 2019; Monteiro et al., 2021; Hoai and Nguyen, 2022; Tetteh et al., 2022; Nguyen et al., 2023).

Here the departure lies not in reporting on relationships but in reporting certain things concerning the measure itself. According to these authors, they describe their measures as organizational attributes. But according to the findings reported here, at least in this kind of situation, such measures appear mainly to be respondent-level phenomena whose status as organization-level indicators was never demonstrated because no one ever calculated the intraclass correlations necessary for making that judgment. Hence the claim made does not rule out everything said before but instead adds something to what had previously been stated.

This point relates once again to accountability in MFIs (micro-finance institutions). According to Akanga (2017) and Marini et al. (2018), there are accountability practices that fulfill external requirements but actually tell nothing about their addressees. Hence, if perceptual measures reflect more about respondents themselves than institutions, then they will show precisely this kind of formal compliance without substance; here again, some indication of its size appears from these findings.

5.3 Implications for Research Design and Reporting

Interpretation: A one-informant measure should be called precisely that, rather than something else; one informant's judgment does correlate reasonably well but still only moderately with what others who know this person would say. Findings concerning organizations based on these measures are therefore much less certain than usually stated, especially if they involve ranking different institutions.

As regards the number of informants, the answer depends on the goal. For a study concerned with relationships among individuals' own reports, one informant from every organization will suffice provided that the cluster structure is reported. Where the aim concerns some quality attributable to organizations themselves, however, there need to be enough informants to measure it. At the intraclass correlation levels found above, obtaining reliability of 0.70 for an institution mean requires approximately fifteen cases per institution. This figure exceeds many investigators' resources altogether, which makes it another very significant result concerning what such measurement can achieve.

For reporting, intraclass correlations must also appear as a matter of routine because, where an institution identifier exists, there is no extra expense; readers then learn what the magnitudes of the standard errors mean, whereas otherwise nobody knows how precise the total available evidence is.

6.0 CONCLUSION AND IMPLICATIONS

6.1 Summary of Findings

This study investigated how much of a perceptual measure of internal control quality reflects the institution being described, and whether the two vantage points within an institution describe the same thing, drawing on 320 key informants in 80 Category 1 microfinance institutions in Cameroon. It produced two principal findings.

Asking just one person about an organization produces quite a poor measure, insofar as the organization makes up barely more than one-third of what is measured: some fourteen percent of perceived internal-control quality here, with $r = .59$ when someone correlates with the average rating made by everyone else working there, and even the mean of four people still attaining only .39 reliability. This need not invalidate such questions altogether, but it again means they contain rather little organizational information despite how most often described.

And secondly, managers and operators have the same views regarding how much control there is overall and concerning its connection with the correctness of the information available, but they rank individual agencies quite differently, hence their mean scores correlate very poorly (only $r = +.34$). Hence also, whose opinion this actually reflects makes no difference whatsoever if researchers wish merely to learn something true about everyone collectively, yet matters greatly when they seek insight into specific bodies themselves; neither result could possibly emerge unless investigators record alongside every answer given by whom precisely at that place, and then question people holding these two positions everywhere accordingly.

6.2 Practical Implications

Here again, there can be no recommendations for researchers that are too particular or inexpensive: record an institution identifier on each questionnaire; when the object being studied concerns organizations themselves, obtain information from more than one person per organization; report ICC(1) and ICC(2) along with standard indicators of reliability; otherwise, one cannot even determine whether something was measured at the level of individuals or institutions.

The lesson about order also applies directly if an umbrella body itself monitors its member organizations' controls and obtains members' reports. When control quality depends entirely upon how some officer feels about his own unit, then what members report collectively determines their relative rank order among themselves. But supervisors who distribute resources according to these ratings, or decide whether to inspect anyone according to this index, cannot rely on just one person's opinion throughout their entire organization; they need additional proof beyond people's feelings alone.

Regulators and their developers can take some comfort here: sector-level pictures derived from perceptions appear much more credible than institution-level portraits produced similarly; statements regarding what proportion of Category 1 bodies have satisfactory IC overall will therefore be better founded than choices concerning whom to approach first.

6.3 Limitations and Future Research

With four informants per institution, estimating what each institution's mean actually is itself involves error, which further reduces correlations among vantage points and makes the observed level of consensus lower than its true value. It is also a snapshot, saying nothing about whether shared understanding grows over time if people spend more periods working together. Its location limits generalization: although the theory carries over, there is no reason why precisely these figures concerning how much overlap exists ought to hold elsewhere. Finally, everything concerns what individuals believe rather than an actual objective standard by reference to which someone might later determine whose reports were correct on their respective topics.

Three directions seem worth following here. One concerns how one designs the control-system measurement itself in order to distinguish measurement noise from real within-institution differences (something this survey-style work cannot achieve); another involves taking multiple observations longitudinally to find out whether there is convergence over time as people spend more time together or whether agreement is constant; finally, someone needs to link these kinds of reports to outside indices because none exists now.

REFERENCES

- Aguirre-Urreta, M. I., and Hu, J. (2019), Detecting common method bias: Performance of the Harman's single-factor test. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, 50(2), 45–70. <https://doi.org/10.1145/3330472.3330477>
- Akanga, F. K. (2017), Microfinance accountability in Cameroon: A cure or a curse for poverty alleviation? *Journal of Accounting & Organizational Change*, 13(1), 112–130. <https://doi.org/10.1108/JAOC-11-2015-0087>
- Chalmers, K., Hay, D., and Khelif, H. (2019), Internal control in accounting research: A review. *Journal of Accounting Literature*, 42, 80–103. <https://doi.org/10.1016/j.acclit.2018.03.002>
- Chin, W. W., Thatcher, J. B., and Wright, R. T. (2012), Assessing common method bias: Problems with the ULMC technique. *MIS Quarterly*, 36(3), 1003–1019. <https://doi.org/10.2307/41703491>
- Committee of Sponsoring Organizations of the Treadway Commission. (2013), Internal control—Integrated framework. AICPA.
- Cornforth, C. (2004), The governance of cooperatives and mutual associations: A paradox perspective. *Annals of Public and Cooperative Economics*, 75(1), 11–32. <https://doi.org/10.1111/j.1467-8292.2004.00241.x>
- Dijkstra, T. K., and Henseler, J. (2015), Consistent partial least squares path modeling. *MIS Quarterly*, 39(2), 297–316. <https://doi.org/10.25300/MISQ/2015/39.2.02>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., and Sarstedt, M. (2022), A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). Sage.
- Hair, J. F., Risher, J. J., Sarstedt, M., and Ringle, C. M. (2019), When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Henseler, J., Ringle, C. M., and Sarstedt, M. (2015), A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hermes, N., and Hudon, M. (2018), Determinants of the performance of microfinance institutions: A systematic review. *Journal of Economic Surveys*, 32(5), 1483–1513. <https://doi.org/10.1111/joes.12290>
- Hoai, T. T., and Nguyen, N. P. (2022), Internal control systems and performance of emerging market firms: The moderating roles of leadership consistency and quality. *SAGE Open*, 12(3). <https://doi.org/10.1177/21582440221123509>
- Kock, N. (2015), Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10. <https://doi.org/10.4018/ijec.2015100101>

- Liang, H., Saraf, N., Hu, Q., and Xue, Y. (2007), Assimilation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quarterly*, 31(1), 59–87. <https://doi.org/10.2307/25148781>
- Marini, L., Andrew, J., and van der Laan, S. (2018), Accountability practices in microfinance: Cultural translation and the role of intermediaries. *Accounting, Auditing & Accountability Journal*, 31(7), 1904–1931. <https://doi.org/10.1108/AAAJ-07-2017-3028>
- Mersland, R., and Strøm, R. Ø. (2009), Performance and governance in microfinance institutions. *Journal of Banking & Finance*, 33(4), 662–669. <https://doi.org/10.1016/j.jbankfin.2008.11.009>
- Monteiro, A. P., Vale, J., Silva, A., and Pereira, C. (2021), Impact of the internal control and accounting systems on the financial information usefulness: The role of the financial information quality. *Academy of Strategic Management Journal*, 20(Special Issue 3), 1–13.
- Nguyen, T. N. L., Vu, T. K. A., and Bui, T. N. (2023), Impact of internal control on the quality of accounting information at enterprises listed on the stock market in Vietnam. *International Journal of Professional Business Review*, 8(5), Article e01010.
- Nunnally, J. C., and Bernstein, I. H. (1994), Psychometric theory (3rd ed.). McGraw-Hill.
- Omidiji, A., Redmayne, N. B., Ehalaiye, D., and Gyapong, E. (2024), Internal audit in microfinance institutions: Evidence from transitional and developing economies. *Economics of Transition and Institutional Change*, 32(1), 109–141. <https://doi.org/10.1111/ecot.12382>
- Phornlaphatrachakorn, K. (2019), Internal control quality, accounting information usefulness, regulation compliance, and decision-making success: Evidence from canned and processed foods businesses in Thailand. *International Journal of Business*, 24(2), 198–215.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003), Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879–903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Podsakoff, P. M., Podsakoff, N. P., Williams, L. J., Huang, C., and Yang, J. (2024), Common method bias: It's bad, it's complex, it's widespread, and it's not easy to fix. *Annual Review of Organizational Psychology and Organizational Behavior*, 11, 17–61. <https://doi.org/10.1146/annurev-orgpsych-110721-040030>
- Ringle, C. M., Wende, S., and Becker, J.-M. (2024), SmartPLS 4 [Computer software]. SmartPLS GmbH. <https://www.smartpls.com>
- Sarstedt, M., Hair, J. F., Cheah, J.-H., Becker, J.-M., and Ringle, C. M. (2019), How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal*, 27(3), 197–211. <https://doi.org/10.1016/j.ausmj.2019.05.003>
- Sarstedt, M., Hair, J. F., Pick, M., Lienggaard, B. D., Radomir, L., and Ringle, C. M. (2022), Progress in partial least squares structural equation modeling use in marketing research in the last decade. *Psychology & Marketing*, 39(5), 1035–1064. <https://doi.org/10.1002/mar.21640>

- Streukens, S., and Leroi-Werelds, S. (2016), Bootstrapping and PLS-SEM: A step-by-step guide to get more out of your bootstrap results. *European Management Journal*, 34(6), 618–632. <https://doi.org/10.1016/j.emj.2016.06.003>
- Tadele, H., Roberts, H., and Whiting, R. H. (2022), Microfinance institutions' risk and governance in Sub-Saharan Africa. *International Journal of Social Economics*, 49(3), 449–469. <https://doi.org/10.1108/IJSE-10-2020-0719>
- Tetteh, L. A., Kwarteng, A., Aveh, F. K., Dadzie, S. A., and Asante-Darko, D. (2022), The impact of internal control systems on corporate performance among listed firms in Ghana: The moderating role of information technology. *Journal of African Business*, 23(1), 104–125. <https://doi.org/10.1080/15228916.2020.1826851>
- van Beest, F., Braam, G., and Boelens, S. (2009), Quality of financial reporting: Measuring qualitative characteristics (NiCE Working Paper 09-108). Nijmegen Center for Economics, Radboud University Nijmegen.
- Willis, G. B., Cognitive Interviewing: A Tool for Improving Questionnaire Design. 2005. Thousand Oaks, CA: Sage.