USAID
FROM THE AMERICAN PEOPLE

# ENHANCED PRINDEX APPLICATION IN COLOMBIA

## ASSESSMENT OF ALTERNATIVE SURVEY INSTRUMENTS FOR MEASURING PERCEIVED TENURE SECURITY FOR LAND AND PROPERTY RIGHTS

### INTEGRATED LAND AND RESOURCE GOVERNANCE TASK ORDER UNDER THE STRENGTHENING TENURE AND RESOURCE RIGHTS II (STARR II) IDIQ

Contract Number: 7200AA18D00003/7200AA18F00015
COR: Sarah Lowery
USAID Office of Land and Urban
Contractor Name: Tetra Tech
Author(s): Denys Nizalov, Adriana Gaviria, Malcolm Childress, and David Varela (Global Land Alliance)

**DECEMBER 2022**

This document was produced for review by the United States Agency for International Development. It was prepared with support from the Integrated Land and Resource Governance Task Order, under the Strengthening Tenure and Resource Rights II (STARR II) IDIQ. It was prepared by Global Land Alliance under subcontract to Tetra Tech.att

**ACKNOWLEDGEMENTS**

Funding for the Enhanced Prindex Application in Colombia – Design, Coordination and Analysis was made possible by the United States Agency for International Development (USAID) under the Strengthening Tenure and Resource Rights II (STARR II) IDIQ through USAID Contract No. 7200AA18D00003/7200AA18F00015 (USAID's Integrated Land and Resource Governance). The technical support and operational leadership for data collection was provided by the Colombian data collection firm, G|Exponencial and by and the broader Prindex research team. Mónica Ribadeneira provided valuable inputs for the linguistic analysis. Kevin Barthel and Elyse Magen provided valuable inputs and support through all stages of project implementation. The close collaboration of these partners permitted the Prindex survey to be completed, despite the security and COVID-19 challenges present in Colombia during the study period. Global Land Alliance is grateful for the collaboration and continual support that made this report possible. The team greatly appreciates comments on the project concept and on the earlier version of this report provided by Caleb Stevens, Karol Boudreaux, Gregory Haugan, Heather Huntington, Benjamin Linkow, Lauren Persha, Ali Protik, Ximena Rueda, and Jolyne Sanjak.

Cover Photo: G|Exponencial survey enumerator collecting data in San Juan Nepomuceño. Elyse Magen/Global Land Alliance

Tetra Tech Contact: Megan Huth, Project Manager
159 Bank Street, Suite 300
Burlington, VT 05401
Tel: (802) 495-0282
Fax: (802) 658-4247
Email: megan.huth@tetratech.com

Suggested Citation: Nizalov, D., Gaviria, A., Childress, M., & Varela, D. (2022). *Enhanced Prindex application in Colombia: Assessment of alternative survey instruments for measuring perceived tenure security for land and property rights*. Washington, DC: USAID Integrated Land and Resource Governance Task Order under the Strengthening Tenure and Resource Rights II (STARR II) IDIQ.

# ENHANCED PRINDEX APPLICATION IN COLOMBIA

## ASSESSMENT OF ALTERNATIVE SURVEY INSTRUMENTS FOR MEASURING PERCEIVED TENURE SECURITY FOR LAND AND PROPERTY RIGHTS

### INTEGRATED LAND AND RESOURCE GOVERNANCE TASK ORDER UNDER THE STRENGTHENING TENURE AND RESOURCE RIGHTS II (STARR II) IDIQ

Submission Date: 3 October 2022
Submitted by: Melissa Hall, Deputy Chief of Party
Tetra Tech
159 Bank Street, Suite 300, Burlington VT 05401, USA
Tel: (802) 495-0282
Fax: (802) 658-4247

Contract Number: 7200AA18D00003/7200AA18F00015
COR Name: Sarah Lowery
USAID Office of Land and Urban
Contractor Name: Tetra Tech
Author(s): Denys Nizalov, Adriana Gaviria, Malcolm Childress, and David Varela (Global Land Alliance)

# TABLE OF CONTENTS

# LIST OF ACRONYMS

| | |
|---|---|
| AIC | Akaike Information Criterion |
| BIC | Bayesian Information Criterion |
| DANE | Departamento Administrativo Nacional de Estadística/National Administrative Department of Statistics |
| DK/NR | Don't Know/No Response |
| FAO | Food and Agriculture Organization of the United Nations |
| GSEM | Generalized Structural Equation Model |
| IC | Information Criterion |
| IDIQ | Indefinite Delivery/Indefinite Quantity |
| IE | Impact Evaluation |
| ILRG | Integrated Land and Resource Governance Program |
| LFP | Land for Prosperity Activity |
| LL | Log-Likelihood Measure |
| MTMM | Multitrait-Multimethod methodology |
| NORC | National Opinion Research Center at the University of Chicago |
| OBS | Observations |
| OECD | Organisation for Economic Co-operation and Development |
| OR | Odds Ratio |
| PTS | Perception of Tenure Security |
| SB-MTMM | Split-Ballot Multitrait-Multimethod |
| SDG | Sustainable Development Goal |
| SQP | Survey Quality Predictor |
| STARR II | Strengthening Tenure and Resource Rights II |
| TPTS | "True" Perception of Tenure Security |
| UN | United Nations |
| UN-Habitat | United Nations Human Settlement Programme |
| USAID | United States Agency for International Development |

# EXECUTIVE SUMMARY

The research was designed and executed under the United States Agency for International Development's (USAID's) Integrated Land and Resource Governance (ILRG) activity. Data collection was carried out from May through July 2021.

Secure and transferable property rights for land and housing are a key driver of economic development, environmental sustainability, and social stability, and are an important focus of public policy. To better protect property rights and develop effective policies and programs in this area (e.g., the USAID Land for Prosperity Activity [LFP] in Colombia), policymakers and practitioners require a clear picture of the current level of tenure security[1] across countries, sub-national jurisdictions, and groups of people, and they need to be able to track changes across time. An increasingly popular measure of tenure security is individuals' perceptions of tenure security (PTS) – a subjective assessment of the risk of losing property rights. The main benefit of PTS over other measures[2] of tenure security, such as possession of formal/government-issued documents confirming property rights, is related to its comparability across tenure forms (including informal) and legal systems. However, as has been noted in survey design literature and program implementation practice, differences in how to measure PTS may lead to inconsistencies in assessment and thus in programmatic and policy recommendations.

This study draws on new data collected in Colombia of individuals' PTS, based on the Prindex methodology, to learn more about the formulation of measurement scales[3] for this increasingly used metric. It is motivated by the expanding set of studies of tenure security around the world using different measurement scales and a corresponding need to understand the impact of these differences for comparability and reliability of the results used in policy discourse. This study also contributes to learning about tenure security in some of the most conflict-affected areas of Colombia. The data from Colombia was collected using two different measurement scales and allows analysis of the impact of these approaches. It provides corresponding recommendations for this type of research.

With the goal of reducing inconsistencies across and improving the use of PTS measures to inform policies, program designs and results monitoring, this study compares two widely used alternative PTS response scales:

1. A **4-point symmetrical scale** (*very unlikely, unlikely, somewhat likely, very likely*; a scale with an equal number of positive and negative categories) used by Prindex in 140 countries; and
2. A **5-point asymmetrical scale** (*not at all likely, slightly likely, moderately likely, very likely, extremely likely*; a scale with an unequal number of positive and negative categories) proposed by the Food and Agriculture Organization of the United Nations (FAO), World Bank, and the United Nations

[1] This study is based on the Prindex methodology (https://www.prindex.net). In Prindex, property right refers to a bundle of rights to access, use, possess/control, and benefit from housing and land. Different tenure types, such as leasehold/use rights or ownership/allodial rights, consist of different bundles of rights. Prindex methodology allows measurement of perceived tenure security for tenure types and for rights which are formal or informal, individual or group, and customary or statutory in nature. This broad definition of property rights is more commonly referred to as land tenure, and tenure security correspondingly is a core quality of property rights having to do with their enforceability in practice.

[2] The term "measure" is used to refer to a specific representation of the general concept of tenure security using data and a specific methodology of constructing such a representation out of available data (e.g., defining what represents secure or insecure tenure based on respondent's answers).

[3] The term "measurement scale" refers to the way survey response options are characterized, an element of survey question design, to collect data on tenure security to be used to construct measures of tenure security.

Human Settlement Programme (UN-Habitat) in their recommendations on data collection for the Sustainable Development Goal (SDG) indicators.[4]

The new dataset produced for this study is custom designed to generate an "experiment" that compares these two scales as well as several alternative measures of PTS in a homogeneous population in rural Colombia. Both measurement scales are attempting to capture a respondent's real, unobservable level of tenure security, the degree to which they truly feel secure or insecure about their property rights. This is referred to as their "true perception of tenure security" (TPTS). Yet numerous studies suggest that there is often a disconnect between a respondent's internal assessment of their level of tenure security and how they report their perception of security on a given measurement scale. This disconnect affects the quality of the data. Thus, variation in responses between two respondents could reflect variations in the underlying level of TPTS, systematic measurement error associated with a particular survey instrument, or random error. In order to assess the data quality (measured as reliability and validity) of the two scales, the study uses a regression analysis (specifically a method called path analysis) to develop a proxy measure, or predicted value, of TPTS, based on eleven alternative questions related to PTS asked in the survey. The correlation between the predicted value of TPTS and the values given on the two measurement scales is used as an estimate of data quality for each instrument. The stronger the correlation, the higher the data quality.

According to this comparison, the 4-point and 5-point scales both display advantages. The 4-point scale is easy to interpret and translate into other languages, given the equal number of positive and negative categories of tenure security. This may allow for easier comparison across countries. The 5-point scale allows for more nuanced measures of tenure security and may be better able to pick up small shifts in security over time. As a result, it may be better suited for impact evaluations that are trying to accurately capture small changes as the result of program interventions. Yet neither scale is perfect. The linguistic precision of the terms used in the response options (especially when translating from English to locally spoken languages), the respondents' understanding of the response options, and how the mid-point response option in the 5-point measurement scale is interpreted may lead to misrepresentation of the actual level of tenure security.

Given the advantages and disadvantages detected in both measurement scales, the study concludes that converting results from either of the compared scales into a binary variable (secure tenure vs. insecure tenure) yields more reliable results than either of the two unadjusted alternatives for cross-country comparison. The study identifies a need for some adjustments which can be made to both of the scales to improve the accuracy of such conversion. This analysis suggests that somewhat more adjustment is needed to the 5-point asymmetrical scale currently used for collecting SDG official data than for the 4-point symmetrical scale used currently by Prindex. A focus on improving the scales and a shift to reporting on binary converted results would allow aggregation of data across sources, expanding the size of comparable datasets available to improve our understanding of PTS.

More specifically, the key conclusions of the experiment using ILRG-collected data in Colombia include:

- **Number of categories:** The is no one correct way to classify "true" (predicted) PTS. In fact, different sub-populations (e.g., municipalities, holders of different forms of tenure) can be optimally classified into four, five, or six categories. The optimal number of categories for a particular data set depends on the proportions of the relevant sub-populations in the data, which is not known to a researcher prior to data collection. For consistency and comparability across countries and institutional contexts, either 4- or 5-point scales can be selected as a

[4] Annex 3 additionally provides some comparison of the 4-point symmetrical scale with a 5-point symmetrical scale that is in use in USAID impact evaluations. This comparison, however, is not the main focus of this ILRG Prindex study.

standard and consistently applied. In cases when comparability is less important, the practicality of scale alternatives in a particular context should be considered on a case-by-case basis.

- **Labelling and data quality**: The shorter 4-point scale assessed in this study may be easier to interpret and implement across different languages and cultural environments (e.g., in relation to preserving the relative order of categories, their interpretation as positive or negative, and a relative distance among the categories in terms of the level of tenure security):
    - The asymmetrical nature of the 5-point scale assessed in this study and, more generally, the use of a larger number of categories, requires application of more complex language in the response options, to describe the nuanced differences in tenure security across the larger number of categories. The analysis suggests that more complex language can be more difficult for respondents to interpret in comparable ways.

      For example, a comparison of the asymmetrical 5-point scale with an alternative symmetrical 5-point scale (two negative categories, two positive categories, and a neutral middle category; used for data collection for the LFP impact evaluation baseline conducted by the National Opinion Research Center at the University of Chicago [NORC]) covering the same municipalities demonstrates a very different distribution of responses across the 5-point response categories (see Annex 3).
    - In contrast, the 4-point scale is better able to overcome the above difficulties as a shorter scale can rely on simpler language to describe the categories, which is reflected in the lower percentage of misclassified observations.

      This advantage is demonstrated in the comparison of the observed versus predicted values of PTS for each scale. About 89 percent of observations of PTS measured with the 4-point scale matched the categories of predicted PTS. In contrast, the 5-point scale classified only 34 percent of observations correctly across the five categories of predicted PTS.
- **Matching the target population:** As noted, the 5-point scale uses language of higher complexity to express nuances in meaning among a larger number of categories, which may not match the levels of language typically used by the survey respondents. For example, some feedback from survey enumerators was that that the labels were "too academic," which is confirmed by the linguistic analysis conducted in this study. A shorter scale requires less complex language and provides a better match across languages and linguistic skill levels. Survey designers should consider the educational level of the target population when deciding between various measurement scales, alongside robust pilot testing of survey instruments.
- **Middle point interpretation:** The 5-point scale is more vulnerable to errors in converting the data on PTS into a comparable binary measure (secure/insecure). In particular, the presence of a middle point in the asymmetrical 5-point scale creates a risk of misclassification, as the data analyst's allocation of the middle category as "secure" or "insecure" may not match the respondent's "true" perception of tenure security. The 4-point symmetrical scale minimizes this risk. The respondent must make an explicit choice between the secure and insecure categories without any assumptions by the data analyst. This conclusion is specific to PTS measures that ask about the likelihood of losing property rights, as there is no natural "middle" category to explain tenure security.
- **Comparability:** Reclassification of data into binary outcomes (secure/insecure) reduces misclassification error that can exist among the more granular categories and improves comparability between PTS measures based on 4-point and 5-point scales. However, the comparability of the binary outcomes created through reclassification is conditional on the

correct allocation of the middle points on the 5-point scale into either the secure or insecure category. One has to assume that all or some specific proportion of the respondents who selected the middle point should be treated as secure. This may be higher or lower than the true level of tenure security in one environment (biased). However, the true level of tenure security among such respondents may be different in another environment (inconsistent).

This study also provides a snapshot of findings about PTS itself (versus methodological findings on the alternative measurement scales) for the municipalities where data was collected for this experiment. It provides for comparison to findings on PTS from monitoring data also collected in 2021 for massive formalization pilots under LFP, as well as a 2018 Prindex national poll in Colombia (Table 1 below).[5]

**TABLE 1. PERCENTAGE OF ADULTS THAT PERCEIVE THEIR TENURE AS INSECURE BY GENDER, GEOGRAPHY, AND TENURE TYPE**

| Sample | Overall + | Gender | | Geography | | Tenure Type | | | | Type of Threat | | Lost rights in the past |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Male | Female | Urban | Rural | Own | Rent | Family | Other | Internal to Family | External Actors | |
| LFP comparison (ILRG) | 32% | 33% | 33% | 35% | 30% | 16% | 55% | 35% | 52% | 10% | 29% | 38% |
| LFP pilot monitoring | 33% | 31% | 34% | 32% | 34% | 23% | 58% | 30% | 53% | 13% | 23% | 33% |
| Prindex national poll | 24% | 24% | 23% | 23% | 23% | 10% | 31% | 24% | 31% | 5% | 15% | 9% |

Prindex measures percent of the adult population that feels secure or insecure in their property rights for the main housing property they reside in and its attached land

Per Table 1, the datasets show almost identical levels of tenure insecurity at 32 percent of respondents in the experiment's dataset and 33 percent of respondents in the LFP pilot dataset. Although the ILRG and LFP-funded surveys were developed for different purposes and carried out in two similar but distinct sets of rural municipalities in Colombia, the key results are strongly similar in both surveys. The comparison tends to reinforce the validity of the results of the experiment, as the results from the LFP pilot data (non-experimental, 4-point symmetric scale only) record similar basic parameters in the key variables of interest.

The levels of insecurity observed in the ILRG and LFP datasets are eight and nine percent higher than the 2018 national survey average of 24 percent, a difference likely due to the higher non-response rate in the national poll (the reduction in non-response in the ILRG and LFP datasets is likely due to improved interviewer training). Table 1 also indicates that the differences in PTS between the national level data and the ILRG and LFP data, representative of the municipalities selected, are similar across gender and urban versus rural areas. Owners and renters in the LFP and ILRG datasets, however, reported proportionately more perceived tenure insecurity than in the national survey, likely because the municipalities are in areas that were highly affected by Colombia's civil war and consistent with the higher percentage of respondents that reported having lost property in the past.

Perceptions of tenure security are also relevant for people's access to land-based resources (e.g., pastures, water) and infrastructure (e.g., roads), and in this study we added questions about these topics. The results highlight this as another area to learn more about. For example, 14 percent of the respondents in the ILRG dataset report feeling that it is likely they may lose their access to water

[5] The results found in the LFP pilot monitoring dataset are found in Barthel et al., 2021; the 2018 Prindex national poll results are found in Feyertag et al., 2020.

resources in the next five years. This could be an important topic for learning in relation to climate change adaptation and sustainable food systems and food security.

While this study is based on data from a particular context – Colombian municipalities affected by recent violent conflicts – and the versions of the 4- and 5-point scales used were translated only in one language, Spanish, the main conclusions of the experiment (above) shed light on how differences in measurement scales can affect our ability to consistently and comparably measure PTS as needed for evidence-based policy and program adaptations. At the same time, additional research is needed to refine this learning both contextually and methodologically.

To conclude, three practical implications of this study for researchers and policymakers are:

1. A binary measure of tenure security should be constructed for comparison across countries (of comparable forms of tenure and types of property), regardless of the design of original scale used for the data collection (OECD, 2013). Particular attention is needed to ensure clarity and consistency in constructing the binary measure, as the original question/measurement scale used affects the quality of the binary outcome. The data collection tool itself should use a measurement scale with more response options (e.g., 4- or 5-point scales, depending on context). The reclassification into a binary measure should occur after data has been collected.
2. On the other hand, the gain in comparability comes at a cost of losing some nuance and information on the true variation in the levels of tenure security. Such information may be particularly useful when measurement needs to be more sensitive to changes in PTS that are relatively small in magnitude or affect only a part of the distribution (e.g., people were already secure, but have become even more confident in their tenure security over time).
    - A shorter (4-point) symmetrical scale demonstrates a higher robustness to misclassification and errors in translation and is easier for implementation and conversion into a binary scale. Therefore, it is recommended for assessment of PTS across countries and institutional environments.
    - The 5-point scale may be more practical in cases where the measurement needs to be more sensitive to small changes in PTS (e.g., in impact evaluations). However, the design of labels for a 5-point scale needs more consideration. A scale with a larger number of positive (secure) categories than negative is likely to demonstrate better data quality results.
3. Additional studies are necessary to check the robustness of these recommendations to other languages and institutional contexts. Consistency in the global assessment of tenure security will be facilitated by the development and testing of standard labels for tenure security scales in multiple languages with a clear guidance for categorization into secure or insecure tenure.[6]

[6] Publication of Prindex instruments used in 140 countries in local languages could be a starting point in this process since a large investment has already been made into the initial set of translations.

# 1.0 INTRODUCTION

This report presents the results of an experiment measuring perception of tenure security (PTS) for land and housing property using alternative measurement scales in four municipalities of Colombia. This tenure security assessment is based on the Prindex methodology[7], which measures PTS in a comparable way across institutional settings and individual tenure arrangements, as described below. This methodology and questionnaire were modified to support the experimental comparison.

The experiment was conceived and executed as an activity under the United States Agency for International Development's (USAID) Integrated Land and Resource Governance (ILRG) program. The survey data collection was carried out from May through July 2021.

[7] Information about the Prindex initiative can be found on its website (https://www.prindex.net).

# 2.0 BACKGROUND

Secure and transferable property rights for land and housing are a key driver of economic development, environmental sustainability, and social stability, and are an important focus of public policy. They can facilitate investment, improve productivity, reduce poverty, help build smart and inclusive cities, as well as improve inclusivity and justice (see Besley, 1995; Deininger, 2003; Lawry et al., 2017). To deliver improved protection of property rights, policymakers and practitioners need a clear picture of the current level of tenure security[8] across countries, sub-national jurisdictions, and groups of people, and to be able to track changes across time. This can help identify the most vulnerable groups and the effectiveness of land governance policy and innovations across countries, regions, and municipalities. While it is now understood that documented land rights can increase tenure security, tracking progress by tracking changes in the extent of documentation is a traditionally used but imperfect proxy. The real or perceived risk of losing one's rights to property a) doesn't only depend on documentation and b) affects behavior. Therefore, an increasingly popular measure of tenure security is an individual's perceived tenure security (PTS) – a subjective risk assessment of the likelihood of losing one's property rights. PTS is now being used in many academic research and policy projects including reporting on Sustainable Development Goal (SDG) 1.4.2 – the proportion of the total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure, by sex and type of tenure. Another advantage of PTS over other measures[9] of tenure security, such as possession of formal documents confirming property rights, is its comparability across tenure forms (including informal) and legal systems. However, as noted in the survey design literature and program implementation practice, differences in how to measure PTS may lead to inconsistencies in assessment and thus in programmatic and policy recommendations.

This study draws on new data collected in Colombia of individuals' perceptions of their security of property rights, or tenure security, based on the Prindex methodology. The goal was to learn more about the formulation of measurement scales[10] for this increasingly used metric. It is motivated by the expanding set of studies of tenure security around the world that use different measurement scales and the need to understand the impact of these different scales on the robustness of results so that conclusions of these studies can become more comparable and reliable in policy discourse. The new data from Colombia was collected using two different measurement scales (the survey design is described in Section 4), which allows analysis of impact of these different scales and corresponding recommendations for additional research. This specialized focus on the measurement scales is justified by the wide and generalized economic and social importance of the theme which they measure. At the same time, this study contributes to learning about tenure security in some of the most conflict-affected areas of Colombia.

Currently, two survey instruments for measuring PTS are gaining popularity. One is promoted by Prindex and has been used to collect data on PTS across 140 countries (Feyertag et al., 2020). This

[8] This study is based on the Prindex methodology (https://www.prindex.net). In Prindex, property right refers to a bundle of rights of access, use, possession/control, and benefit from housing and land. Different tenure types, such as leasehold/use rights, ownership/allodial rights, consist of different bundles of rights. Prindex methodology allows measurement of perceived tenure security for tenure types and for rights which are formal or informal, individual or group, and customary or statutory in nature. This broad definition of property rights is more commonly referred to as land tenure, and tenure security correspondingly is a core quality of property rights having to do with their enforceability in practice.

[9] The term "measure" is used to refer to a specific representation of the general concept of tenure security using data and a specific methodology of constructing such a representation out of available data (e.g., defining what represents secure or insecure tenure based on respondent's answers).

[10] The term "measurement scale" refers the way survey response options are characterized, an element of survey question design, to collect data on tenure security to be used to construct measures of tenure security.

instrument is based on a subjective assessment of the likelihood of losing property rights in the next five years and uses a 4-point symmetrical measurement scale of response options (see Box 1). The other is an instrument developed by the Food and Agriculture Organization of the United Nations (FAO), World Bank (WB), and United Nations Human Settlement Programme (UN-Habitat) (2019), which is based on a similar assessment question and uses a 5-point asymmetrical[11] measurement scale (see Box 2) and is published as a methodological recommendation to governments reporting on land-related SDG indicators.

**Box 1. Prindex, 2018 questionnaire**

Q46. In the next 5 years, how likely or unlikely is it that you could lose the right to use this property, or part of this property, against your will?

1. Very unlikely
2. Unlikely
3. Somewhat likely
4. Very likely
5. (Don't know)
6. (Refused)

*Options 5 and 6 are in () because they non-substantive and for easier tracking of response rates*

(Prindex, 2018)

There are also slight differences in how the two instruments formulate the question related to the likelihood of losing one's property rights. This reflects the differences in the primary focus of assessment (Prindex: housing with attached land and any other immobile property, FAO/WB/UN-Habitat: agricultural and non-agricultural land plots), the order of the question in the questionnaire (Prindex: Q46, FAO/WB/UN-Habitat: Q13), and in the design of the measurement scale (number of categories on the scale and their labelling).[12] The differences across survey questions have been studied elsewhere.[13] This study compares the two alternative PTS measurement scales, both using the Prindex question wording. It develops practical recommendations for use of each of the measurement scales. The study questions include:

**Box 2. FAO, WB, UN-Habitat question for Non-Agricultural Land***

13. On a scale from 1 to 5, where 1 is not at all likely and 5 is extremely likely, how likely are you to involuntarily lose ownership or use rights to any of the non-agricultural land you own or hold use rights to in the next 5 years?

1. Not at all likely
2. Slightly likely
3. Moderately likely
4. Very likely
5. Extremely likely

* - *A similar question is asked about agricultural land*

(FAO, World Bank, & UN-Habitat, 2019, p. 61)

i. Do the alternative PTS measurement scales provide for a consistent estimate of PTS in a given population?

ii. How do the results collected with alternative measurement scales compare across the populations (samples)?

iii. Which of these two measurement scales provide for a higher quality of data on PTS?

iv. Is the difference in data quality and classification between the alternative measurement scales consistent across gender, urban/rural geographies, forms of tenure, and municipalities?

There are also other instruments for measuring of PTS used by researchers and practitioners, including the 5-point symmetrical scale (with a neutral middle point) used by the National Opinion Research Center at the University of Chicago (NORC), e.g., in their baseline survey for impact evaluation of LFP, and a separate Prindex question based on how "worried" the respondent is about losing the property

[11] By symmetrical and asymmetrical scales, we mean scales that do or do not have an equal number of positive and negative categories offered as response options around the middle point on the scale.

[12] The measurement scale design has implications for what one would consider as secure or insecure tenure. In case of Prindex, the first two categories are treated as secure while prior discussions at workshops and meetings suggest that only one category – not at all likely – is treated as secure in the second case

[13] The effect of wording of the questions is assessed elsewhere (Persha, et al, 2020). The order of the questions is likely to have a relatively small effect (Alwin, 2010) on the data quality.

rights. Such diversity of instruments only highlights the importance of the above questions.[14] The analysis of these questions contributes to learning about how to measure PTS in a more consistent and comparable way, and will motivate additional study on PTS measurement while generating additional evidence on PTS in Colombia. Better PTS measures will help governments and international organizations set priorities for policy and program interventions and assess the effectiveness of such interventions.

There is also a theoretical reason to believe that the differences in measurement scale designs may result in some significant variation in comparability of findings and data quality. Fowler & Fowler (1995) and Saris et al. (2011) argue that the mapping of a personal opinion (e.g., about security of property rights in this study) into the proposed measurement scale may be problematic, as a respondent may not intuit or understand the match between the internal assessment of the unobserved (latent) level of tenure security and what the survey question may propose (Revilla et al., 2014).

Several indicators of data quality for alternative survey instruments are considered in the survey design literature (Alwin, 2010). This analysis relies on the following indicators for comparison of the quality of data collected with the two alternative measurement scales studied:

i. Percentage of non-responses or an item non-response rate;

ii. Reliability and validity, which consider the percentage of variation in the data on PTS explained by the "true" (or, predicted) value of perception of tenure security (TPTS);

iii. Bias in estimated level of PTS (systematic under- or over-reporting compared to the predicted true value of PTS); and,

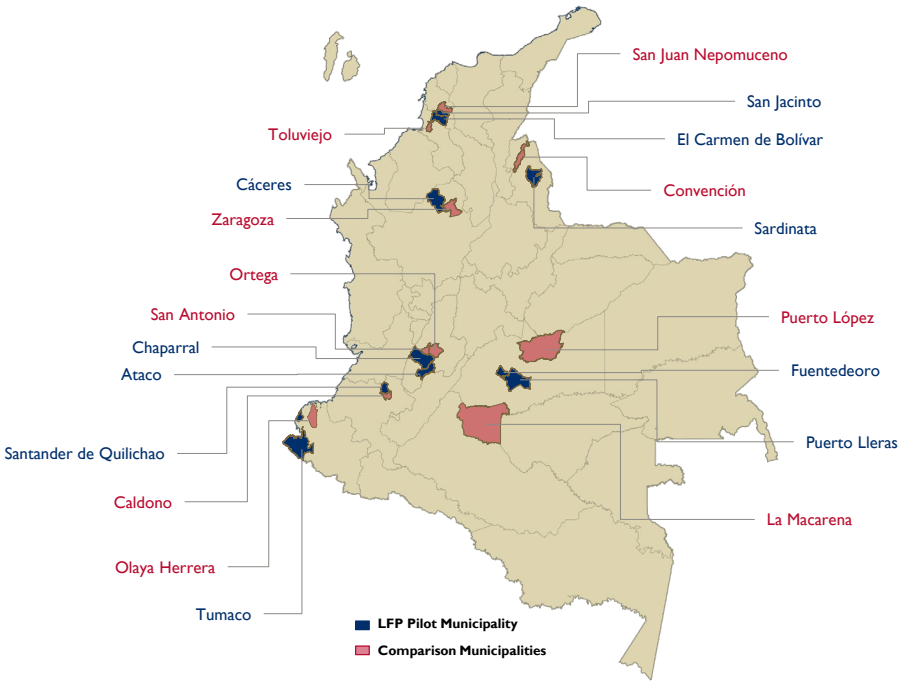
iv. Practical feasibility of implementation and interpretation.

[14] As presented in Table 10 below, several other alternative measures of PTS are used for comparison of data quality to augment learning from this study. Also, some comparison of descriptive statistics provided by NORC is presented in Annex 3. More detailed comparative analyses across alternatives were beyond the scope and budget of this ILRG activity.

# 3.0 A SNAPSHOT OF PRINDEX FINDINGS IN COLOMBIA

The purpose of this section is to provide a synopsis of the findings on perceived tenure security from the data collected under the ILRG-funded Enhanced Prindex Application in Colombia (the "LFP comparison" group of municipalities) with the corresponding findings from other Prindex Colombia data. This will highlight the policy-relevant information gained from measuring PTS before delving into the data analysis related to experimenting with different measurement scales. Findings from the surveys carried out in the four LFP comparison group municipalities (Puerto Lopez, Nepomuceno, Tolu Viejo, and Zaragoza) are compared with the findings from the 10 municipalities included in the LFP pilot surveys[15] ("LFP pilot"). Figure 1 shows both sets of municipalities. The selected municipalities are among those most affected by the armed conflict in the country. Taken together, the LFP pilot and the ILRG-funded surveys have produced one of the largest assessments of tenure security for land and housing in conflict-affected areas globally. The key findings from these recent surveys[16] are also compared to relevant findings of the 2018 national Prindex survey in Colombia. Similarities and differences at the level of descriptive analysis are highlighted along with some specific points of interest. The comparison helps to validate both sets of findings, creating a greater degree of confidence in both datasets. It also helps illustrate where some of the sources of variation in the findings may affect areas of policy interest. This information will be useful for project monitoring, impact evaluation and programming aimed at increasing tenure security throughout Colombia. Finally, the comparative findings support the rationale for focusing efforts to improve tenure security in conflict-affected regions by highlighting differences in findings for these regions compared to broad national averages.

**FIGURE 1. MAP OF COLOMBIA LFP PILOT AND COMPARISON MUNICIPALITIES, 2021**




[15] The LFP pilot survey was conducted between February and April of 2021. Overall, 5,227 individuals participated in the Prindex survey covering the municipalities of Ataco, Caceres, Chaparral, El Carmen de Bolivar, Fuente de Oro, Puente Lleras, San Antonio, San Jacinto, Santander de Quilichao and Tumaco. For more details, please see Barthel et al., 2021.

[16] The ILRG survey data from the sub-sample collected with the 4-point PTS scale was used to construct the descriptive statistics presented in the section so that they are comparable with the other two sources of data presented, which were collected through surveys that used only the 4-point scale and had a comparable sampling strategy.

## 3.1 TENURE SECURITY AND INSECURITY IS SIMILAR ACROSS DATASETS, BY SEX AND BY GEOGRAPHY (RURAL/URBAN)

On average,[17] 65 percent of the adult population in the four LFP comparison municipalities feel secure about their property rights to the main housing in which they reside, and the land attached to it. This is close to the average of the ten LFP pilot municipalities with a tenure security level of 63 percent (Table 2). This proportion is also equal to the national average as recorded by Prindex in 2018 (Feyertag et al., 2020). The results for each dataset are similar across sex (male vs. female) and across geography (rural vs. urban) (Table 3). Such similarity in the level of tenure security tends to co-validate the results of each survey.

While only average results across each dataset are presented in the table, the municipality-by-municipality data draw attention to two municipalities that are outliers with regard to PTS (both from the LFP pilot survey). In the municipality of Cáceres (Department of Antioquia), only 38 percent of adults feel secure about their property rights, which is significantly below the rest of surveyed municipalities. On the other hand, 80 percent of adult residents of the municipality of Ataco (Department of Tolima) feel secure about their rights, which is significantly above the LFP pilot and the national average. The differences in intensity and dynamics of the internal conflict in either municipality may partially explain these results. Cáceres is more exposed to the activities of guerrilla and paramilitary groups and its results may reflect the mid- to long-term impacts of the conflict in terms of perceptions of land tenure security.

Levels of tenure insecurity are also similar across the four LFP comparison and LFP pilot municipalities. About 32 percent of adults in the four LFP comparison municipalities feel insecure about their property rights vs. 33 percent in the LFP pilot municipalities. The 2018 national survey reported 24 percent tenure insecurity. As discussed further below, the results from the two new surveys provide some reason to believe that the level of tenure insecurity recorded in the 2018 national survey may be low due to its higher non-response rate.

**TABLE 2. FINDINGS ON PTS AND RELATED VARIABLES ACROSS, ILRG, LFP, AND PRINDEX NATIONAL DATASETS**

| Survey | % secure+ | % insecure | % non-response | Type of Threat | | | % lost rights in the past | % Possess formal document[18] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Internal | External | Financial | | Total | Owners | Renters |
| LFP comparison | 65% | 32% | 2% | 10% | 29% | 18% | 38% | 31% | 79% | 19% |
| LFP pilot | 63% | 33% | 4% | 13% | 23% | 6% | 33% | 28% | 65% | 24% |
| Prindex national poll | 65% | 24% | 11% | 5% | 15% | 7% | 9% | 57% | 92% | 50% |

Notes: Prindex measures percent of the adult population that feels secure or insecure in their property rights for the main housing property they reside in and its attached land.

[17] For practical purposes and given this section's focus on highlighting differences at the overall sample descriptive statistics across datasets, in this report the computation of averages for the LFP comparison municipalities was done by attributing equal weight to each municipality, regardless of its population size. Findings for each municipality separately are available if LFP should need these and any LFP endline analysis may want to consider including additional weights proportional to the size of the municipality.

[18] Formal documents considered for the purpose of the LFP Prindex baseline survey were title deeds, deeds of contract, survey plans, certificates of customary ownership, certificates of occupancy/possession rights, certificates of hereditary acquisition, registry certificates, sales deeds, registered lease agreements, and rental contracts.

**TABLE 3. PERCENTAGE OF ADULTS THAT PERCEIVE INSECURITY BY GENDER AND GEOGRAPHY**

| Sample | Gender | | Geography | |
|---|---|---|---|---|
| | Male | Female | Urban | Rural |
| LFP comparison | 33% | 33% | 35% | 30% |
| LFP pilot | 31% | 34% | 32% | 34% |
| Prindex national poll | 24% | 23% | 23% | 23% |

## 3.2 NON-RESPONSE IS ASSOCIATED WITH TENURE INSECURITY

In the four LFP comparison municipalities, about two percent of respondents did not provide answers regarding their PTS, which is close to the LFP pilot data (four percent) – both are much lower than the national average of 11 percent. This appears to reflect a higher quality of data collection in the ILRG-funded and LFP pilot surveys, likely due to improved interviewer training since the 2018 national survey.

The decrease in the non-response rate in comparison with the previous studies translates into higher rates of reported insecurity. This finding may suggest that those people who do not answer the tenure security questions are more likely to feel insecure. If this suggestion is true, then the reported levels of insecurity are likely to be on the lower end of the range of the true insecurity value. Although more samples will be needed to confirm this point, the improvement in response rates shows the value of learning from previous rounds to improve data quality and interpretation of older rounds of data.

## 3.3 REASONS FOR TENURE INSECURITY

The reasons for insecurity provide information about the contextual factors that drive tenure insecurity. These reasons are classified as (see Table 2 for statistics):

- Internal to a household:
    - Disagreements with family or relatives;
    - Death of a household member.
- External to household:
    - Owner or renter may require a person to leave;
    - Companies may seize the property;
    - Government may seize the property;
    - Issues with customary authorities.
- Financial:
    - Lack of money or other resources needed to live on the property.

Across the datasets, the most common sources of tenure insecurity are external to households. About 29 percent of adults in the LFP comparison and 23 percent in the LFP pilot municipalities point to such sources. This percentage is higher than the national level of 15 percent; however, it suggests a similar tendency. Among specific reasons given, the threat that the owner or renter may ask the respondents to leave the property is noted most frequently (reported by 22 percent of respondents in the LFP comparison survey, 19 percent in the LFP pilot, 13 percent nationally). The financial reason for

insecurity is the second most frequently cited by the respondents of LFP comparison and the national surveys (reported by 18 percent in the LFP comparison survey, six percent in the LFP pilot, seven percent nationally). Conflict, terrorism, or crime as the reason for insecurity should be highlighted separately. These issues are among the external reasons for insecurity and are cited to a much greater extent in the LFP pilot municipalities (15 percent) and LFP comparison municipalities (nine percent) than the 2018 nation sample average, which was one percent. These differences are likely a result of LFP's focus on municipalities affected by civil conflict and the choice of comparison municipalities with similar characteristics. Once again, Cáceres stands out as a particularly insecure place with more than 60 percent of respondents pointing to conflict, terrorism, or crime as a reason for insecurity.

## 3.4 A PREPONDERANCE OF OWNERS REPORT FORMAL DOCUMENTS

Previous accounts of rural Colombia have estimated that between 20 and 60 percent of land in the country is held under informal arrangements (Barthel et al., 2016). The results from both the LFP and ILRG surveys show that a preponderance of owners possess formal documents supporting their property rights, suggesting that the level of informality among owners is likely closer to the lower end of the range (see Table 2). The four LFP comparison municipalities record a relatively high proportion of owners with formal documents confirming property rights (79 percent in the LFP comparison, 65 percent in the LFP pilot municipalities). In contrast to owners, relatively few renters (19 percent and 24 percent, respectively), or persons living on properties owned by family members (six percent), and under other tenure arrangements (one percent) report having formal documents. Respondents indicated that the lack of formal documents for rental agreements reflects a tendency for such agreements to be based on trust with friends, acquaintances or family members and sometimes can reflect owners not offering to formalize a rental arrangement. The fact that close to 50 percent of people live in family arrangements explains why, overall, only about 28 percent and 31 percent of adult population in the LFP pilot and LFP comparison municipalities, respectively, possess formal documents. This number is much lower than the national average of 57 percent. These statistics suggest there was adequate targeting of the LFP intervention municipalities. However, the most vulnerable municipality of Cáceres (Antioquia) has some of the highest level of possession of formal documents among the LFP municipalities (48 percent), partially as a result of previous formalization efforts in spite of being located in an internal conflict hotspot.

## 3.5 PREVIOUS PROPERTY RIGHTS DISPUTES ARE REPORTED BY A SIGNIFICANT MINORITY OF RESPONDENTS

The survey included a set of questions capturing past experiences of property rights disputes and evictions. The proportion of residents of the LFP pilot municipalities that have ever experienced eviction is more than three times that of the rest of the country (33 percent compared with nine percent). In the LFP comparison municipalities, 38 percent of the adult population have experienced evictions. In Tumaco (Nariño), about 60 percent of respondents have reported evictions in the past, highlighting the level of insecurity in this region of the country that has not been the target of previous formalization programs. Moreover, about 10 percent of evictions in Cáceres (Antioquia) and Chaparral (Tolima) took place over the 12 months prior to the LFP Prindex survey. In addition, 20 percent of respondents in the LFP comparison sample reported having experienced rights' disputes, which is close to the average in the LFP pilot sample (17 percent). Armed conflict is the main reason for previous experience of property rights disputes in both the LFP comparison and LFP pilot municipalities.

## 3.6 STRUCTURE OF TENURE

The difference in sources of insecurity between the LFP pilot, ILRG, and the national survey is driven primarily by the fact that the structure of tenure in the LFP pilot and ILRG municipalities is different than

the national average (Table 4). About 51 percent of adults in the LFP pilot municipalities and 41 percent in the ILRG municipalities are living in and using property that belongs to other family members (compared with 32 percent on average in Colombia). The rental market is underdeveloped with only 11 percent of adults in the LFP pilot and 12 percent in the ILRG municipalities living in non-familial, rented houses (compared with 28 percent on average nationally). Renting is even more insecure than the family arrangements (58 percent of renters in the LFP pilot municipalities and 55 percent in the LFP comparison municipalities report insecure tenure, Table 4). The fact that about 11 percent of respondents in the LFP pilot municipalities (compared with four percent at the national level) report a disagreement with their family members as a reason for insecurity confirms the above statement. Some respondents may be living in family-owned or rented property out of necessity rather than choice (e.g., if they lack affordable and more tenure secure alternatives despite the perceived insecurity of their current arrangements). It may also reflect the results of earlier displacements. The high complexity and cost of estate/probate rules may partially explain this level of uncertainty within families, as well as the variety of living together/marriage/divorce/separation arrangements that are not legally settled.

**TABLE 4. STRUCTURE OF TENURE**

| Sample | Distribution across tenure types | | | | % perceive tenure insecurity | | | |
|---|---|---|---|---|---|---|---|---|
| | Own | Rent | Family | Other | Own | Rent | Family | Other |
| ILRG (LFP comparison) | 34% | 12% | 41% | 13% | 16% | 55% | 35% | 52% |
| LFP pilot | 31% | 11% | 51% | 7% | 23% | 58% | 30% | 53% |
| Prindex national poll | 26% | 28% | 32% | 14% | 10% | 31% | 24% | 31% |

## 3.7 THE ILRG AND LFP SURVEY FINDINGS TEND TO VALIDATE EACH OTHER

Although the two surveys were conceived for different purposes and carried out in two similar but distinct sets of rural municipalities in Colombia, the key results are quite similar in both surveys. This strong similarity of results tends to reinforce the validity of the results of both.

# 4.0 EXPERIMENT DESIGN FOR DATA COLLECTION

The survey team implemented the following experimental design to collect data on PTS in a way that allows for estimation of the difference in data quality associated with the alternative measurement scales studied.

1. The data was collected from a random sample of respondents, representative of the adult population of four municipalities in Colombia (listed in Section 3 above). Each municipality defines a stratum within which a representative sample of rural and urban populations was surveyed (see Table 5 for the number of observations).

2. For each municipality, the sample was split randomly into two sub-samples where two alternative survey instruments were administered: Alternative 1 sub-sample (using the Prindex 4-point scale, Box 1) and Alternative 2 sub-sample (using the FAO/WB 5-point scale, Box 2). The wording of the PTS question, however, stayed the same across the two sub-samples and is based on the Prindex likelihood question presented in Box 1.

3. The Alternative 1 (4-point scale) sub-sample includes 2,275 observations comprising a random sample representative of each municipality's adult population with an over-sampled rural[19] population, sufficient for a margin of error of no more than five percent and adjusted for expected non-response (10 percent in urban areas and up to 25 percent in rural areas). This sampling method is identical to that used in the Prindex survey carried out for LFP in 10 municipalities[20] to establish consistent municipality profiles and baseline measures of program outcomes.

   The Alternative 2 (5-point scale) sub-sample includes 1,759 observations. The sample is representative of each municipality's adult population (without oversampling) and the size is sufficient for achieving a margin of error of no more than five percent for each municipality for the core measure of PTS (based on the "worry" question – $Y_3$, discussed below). The size of the sample was also adjusted for expected non-response (10 percent in urban areas and up to 25 percent in rural areas).

   While most household surveys are conducted with the head of household (which is culturally often a man), the Prindex survey methodology intentionally selects a random sample of respondents from within target area households. This allows the sample to be representative of the adult population in each municipality and allows for the detection of differences across groups such as gender, age, and ethnicity.

[19] For this study we used the *Departamento Administrativo Nacional de Estadística* (DANE) classification of the different types of land in Colombia: *cabecera municipal*, *centro poblado*, and rural dispersed. *Cabecera municipal* is defined as an urban perimeter whose boundaries are established through agreements of the municipal council and where the administrative headquarters of the municipalities are located. In this report we refer to *cabecera municipal* as urban areas. *Centros poblados* are settlements of 20 or more adjacent dwellings organized in an urban manner, with roads and blocks. Rural dispersed corresponds to the areas in which there are no concentrations of construction, and where the uses of land include agriculture, industry, protected natural zones, and areas with no human development such as jungles, natural forests, mangroves, and lakes. As opposed to *centros poblados*, rural disperse dwellings are dispersed and their limits may be defined by natural features (e.g., rivers). Unless stated otherwise, in this report we refer to rural areas as the aggregation of *cabecera municipal* and rural disperse (DANE, 2018).

[20] The ILRG-funded Prindex experiment complements a Prindex assessment of tenure security in 10 municipalities funded by LFP, covering nine pilot municipalities and one comparison municipality. Data collection was done separately for LFP using the same data collection firm for consistency in implementation of the survey. The key findings of the LFP data analysis, as compared with the findings of this experiment, are summarized in Annex 4.

**TABLE 5. SAMPLE SIZE BY MUNICIPALITY**

| Municipality | Department | Urban | *Centro poblado* | Rural dispersed | 4-point sub-sample | 5-point sub-sample | Total |
|---|---|---|---|---|---|---|---|
| Zaragoza | Antioquia | 333 | 155 | 527 | 572 | 443 | 1,015 |
| Puerto Lopez | Meta | 431 | 262 | 320 | 570 | 443 | 1,013 |
| Tolu Viejo | Sucre | 269 | 566 | 176 | 573 | 438 | 1,011 |
| San Juan Nepomuceno | Bolivar | 460 | 421 | 114 | 560 | 435 | 995 |
| | TOTAL | **1,493** | **1,404** | **1,137** | **2,275** | **1,759** | **4,034** |

4. Through administration of the survey instruments, the respondents received either the 4-point scale or the 5-point scale version of the questionnaire. The assignment of the 5-point instrument was pre-programmed with the frequency proportional to the size of the 5-point sub-sample in rural and urban areas of each municipality. The interviewer and respondent were unaware of which group they belong to before beginning administration of the questionnaire (administration of the instruments was blinded at the interviewer/respondent level).

5. Sampling weights were developed to account for probability of selection at each of three stages and adjusted for a systematic non-response at the household level. Ex-post calibration was implemented to fit the key population parameters in urban and rural areas of each municipality.

# 5.0 COMPARISON OF THE EXPERIMENT'S TWO SUB-SAMPLES AND THE QUALITY OF ITS ADMINISTRATION

## 5.1 OVERVIEW

Table 6 presents the descriptive statistics for each of the two sub-samples for selected key variables. It includes the measures of tenure security (Section A); respondent individual characteristics (Section B); tenure status of the primary property (Section C); property characteristics (Section D); selected household characteristics (Section E); and other relevant variables such as proportion of responses from each municipality, share of rural and urban population, and quality of an individual response (Section F). The lack of statistical difference in the above variables between the two sub-samples demonstrates a high quality of administration of the alternative instruments for measuring PTS. The right-most column presents the summary of tests for statistical difference for the complete sample. In 54 out of 55 presented variables no significant difference between the two sub-samples is detected at the 10 percent significance level. In only one case – the prevalence of disputes over property rights – the difference between the two sub-samples is significant at the five percent significance level (with 81 percent of the 4-point sub-sample reporting no dispute experience and 78 percent for the 5-point sub-sample). This variable is excluded from further analysis.

The fact that most variables demonstrate no systematic difference between the two sub-samples suggests that any difference in the measured levels of PTS and the respective data quality between the 4-point scale and 5-point scale are due to the difference in the measurement error (method effect) associated with the respective instruments, rather than with the sub-sample design.

It is also important to comment on the difference in the number of observations among the measures of tenure security presented in Section A. Twelve of 17 presented measures have observations for all respondents in each sub-sample. However, for the remaining five presented measures, fewer respondents replied than were in the total sample. For example, only 1,403 and 999 respondents in each sub-sample respectively responded to the "worry" question about *losing rights in case of divorce*. This non-response rate is related to question construction of this and four subsequent variables, where only a portion of the respondents – in this case, married people – were eligible to respond to the question. Such variables are excluded from further analysis and are presented in Table 6 only as an example of similarity of the two sub-samples.

**TABLE 6. DESCRIPTIVE STATISTICS FOR THE 4- AND 5-POINT SAMPLES**

| Statistics | 4-point sub-sample | 5-point sub-sample | Difference/ significance |
|---|---|---|---|
| Total observations | 2,275 | 1,759 | |
| Margin of error (for secure PTS based on "worry" question) | 3.3 | 3.4 | |
| ***A. Measures of tenure security*** | | | |
| How *worried* are you that you could lose the right to use this property: observation (OBS) | 2,275 | 1,759 | |
| - Secure | 36% | 37% | - |
| - Insecure | 63% | 62% | - |
| - Don't Know/No Response (DK/NR) | 0.3% | 0.3% | - |

| Statistics | 4-point sub-sample | 5-point sub-sample | Difference/ significance |
|---|---|---|---|
| In case of *divorce,* how worried are you that your spouse would have the right to stay but you would be forced to leave this property: OBS | 1,403 | 999 | |
| - Secure | 52% | 52% | - |
| In case your *spouse was to pass away*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 1,403 | 999 | |
| - Secure | 47% | 44% | - |
| In case you *lost your job*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 1,240 | 937 | |
| - Secure | 31% | 29% | - |
| If you had a *disagreement/dispute with your family*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 37% | 34% | - |
| If someone in your *household (besides your spouse) passed away*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 1,329 | 949 | |
| - Secure | 39% | 38% | - |
| If *someone else in your family lost their job*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 38% | 35% | - |
| If you *couldn't make the payments* on this property for two months in a row, how worried would you be that your right to stay in this property would be taken away from you: OBS | 358 | 386 | |
| - Secure | 10% | 9% | - |
| If a *company tried to take over the land* your dwelling is on against your will, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 24% | 25% | - |
| If the *government tried to seize your property* from you (e.g., if they build a road or other infrastructure, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 25% | 25% | - |
| If *another person or group claimed ownership*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 29% | 28% | - |
| If *somebody else fraudulently sells the property*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 25% | 24% | - |
| If a *neighbor initiates a boundary dispute*, how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 43% | 40% | - |
| If a *disagreement arose with local/customary authorities* (e.g., officials/chiefs, elder), how worried would you be that your right to stay in this property would be taken away from you: OBS | 2,275 | 1,759 | |
| - Secure | 34% | 35% | - |
| Did you personally *ever lose the right to live in a property* against your will? OBS | 2,275 | 1,759 | |
| - Secure | 62% | 62% | - |

| Statistics | 4-point sub-sample | 5-point sub-sample | Difference/ significance |
|---|---|---|---|
| Has anyone ever *disputed your right* to live in a property? OBS | 2,275 | 1,759 | |
| - Secure | 81% | 78% | ** |
| In general, how well do you think people in this country are *protected* when it comes to their property rights? | 2,275 | 1,759 | |
| - Secure | 40% | 39% | - |
| ***B. Individual characteristics*** | | | |
| Gender (female) | 50% | 50% | - |
| Age | 43.0 | 43.2 | - |
| Education | | | |
| - Primary or below | 39% | 41% | - |
| - Secondary | 41% | 41% | - |
| - Higher education (technical or university) | 20% | 17% | - |
| Income adequacy | | | |
| - Difficult | 76% | 73% | - |
| - Getting by | 20% | 21% | - |
| - Comfortable | 4% | 5% | - |
| - DK/NR | 0% | 0% | - |
| Marital status (married [formally or cohabitating in an informal union])[21] | 62% | 60% | - |
| Country of birth (Colombia) | 97% | 97% | - |
| ***C. Tenure status of the primary housing property and attached land (mutually exclusive forms of tenure)***[22] | | | |
| Owners/joint owners | 34% | 32% | - |
| Renters/joint renters[23] | 12% | 11% | - |
| Use family property (by members of families of renters or owners who are not paying for the use of property) | 41% | 44% | - |
| Stay with permission | 13% | 13% | - |
| Stay w/out permission | 0% | 0% | - |
| Has other property | 16% | 18% | - |
| ***D. Property characteristics*** | | | |
| Number of rooms | 3.09 | 3.08 | - |
| Wall material (block, brick) | 71% | 72% | - |
| Roof material (metallic or zinc) | 69% | 69% | - |
| Attached to land | 46% | 46% | - |
| Property is used to earn income | 34% | 36% | - |
| Internet: | | | |
| - Yes | 60% | 60% | - |
| - No | 38% | 39% | - |
| - DK/NR | 2% | 1% | - |
| ***E. Household characteristics*** | | | |
| Number of adults | 2.82 | 2.88 | - |

[21] A formal marriage is relatively uncommon in these parts of Colombia. A big share of adult population lives in informal marital arrangements or may have more than one such arrangement.

[22] Analysis of PTS with respect to tenure status is presented in Annex 3.

[23] Q39 of the Prindex (2018) questionnaire defines a renter as a person who makes a payment for use of the property. The payment could be in money or goods and services provided to the owner. A spouse of a renter is considered as a renter.

| Statistics | 4-point sub-sample | 5-point sub-sample | Difference/ significance |
|---|---|---|---|
| Presence of children | 68% | 66% | - |
| ***F. Other*** | | | |
| Urban | 58% | 58% | - |
| *Centro Poblado* | 25% | 25% | - |
| Rural dispersed | 17% | 17% | - |
| Zaragoza | 20% | 20% | - |
| Puerto Lopez | 26% | 26% | - |
| Tolú Viejo | 20% | 20% | - |
| San Juan Nepomuceno | 35% | 35% | - |
| Quality of response: how much influenced by another person | | | |
| - Influence | 87% | 87% | - |
| - No influence | 13% | 13% | - |

*Note: sampling weights are applied; (*) the difference is significant at 10 percent level for the total sample; (**) the difference is significant at five percent level for the total sample; (-) the difference is insignificant at 10 percent level for the total sample.*

## 5.2 COMPARISON OF DESCRIPTIVE STATISTICS ACROSS THE ALTERNATIVE MEASUREMENT SCALES USING THE TWO SUB-SAMPLES OF DATA

Figure 2 and Table 7 present the distribution of responses for each sub-sample of data collected with the two alternative instruments for measuring PTS: with 4-point ($Y_1$) and 5-point ($Y_2$) scales. The histogram (Figure 2) demonstrates some visual similarities in distribution despite the fact that the alternative instruments have a different number of response categories (points on the scale) and that 4-point scale is symmetrical and 5-point scale is asymmetrical. The wording of the category labels is also different.

**TABLE 7. TABULATIONS OF THE ALTERNATIVE SCALES**

| 4-point PTS | Frequency | Percentage | 5-point PTS | Frequency | Percentage |
|---|---|---|---|---|---|
| 1. Very unlikely | 775 | 33.8 | 1. Not at all likely | 570 | 32.0 |
| 2. Unlikely | 735 | 31.1 | 2. Slightly likely | 556 | 32.7 |
| 3. Somewhat likely | 492 | 22.8 | 3. Moderately likely | 355 | 19.5 |
| 4. Very likely | 219 | 10.2 | 4. Very likely | 208 | 12.4 |
| | | | 5. Extremely likely | 39 | 2.0 |
| 5. (Don't know) | 51 | 2.0 | 6. (Don't know) | 29 | 1.4 |
| 6. (Refused) | 3 | 0.1 | 7. (Refused) | 2 | 0.1 |
| **Total** | **2,275** | **100** | | **1,759** | **100** |

*Note: sampling weights are applied*

**FIGURE 2. HISTOGRAMS FOR ALTERNATIVE SCALES**


*Note: sampling weights are applied*

## 5.3 LINGUISTIC COMPARISON OF THE ALTERNATIVE MEASUREMENT SCALES

The purpose of the linguistic comparison is to identify:

- If the wording of labels in the 4- and 5-point measurement scales may affect the quality of comprehension and the quality of mapping unobserved feeling of tenure security into the proposed categories;
- Any specific issues with the language used and lessons learned that can be used for future studies; and
- Recommendations for the consistent conversion of data based on the 4- and 5-point measurement scales into comparable binary measure of tenure security/insecurity (see Section 5.4 for how and why such conversion may be done).

To address the above questions, heuristic analysis by a native Spanish speaker was used as well as a focus group discussion among experienced enumerators and survey designers in Colombia. The following issues were identified:

1. ***Inconsistent translation***. Linguistic communication is a system with structures and combinations that exist in a specific context. It includes codes, signs, and symbols common in a society or particular location. Hence, translations are not always accurate as structures and combinations vary from language to language (Juboori & Muhaisin, 2017). Therefore, translations of survey instruments must be supported by a professional linguist and thorough use of pre-testing.

   Table 8 below presents the English and Spanish versions of the scales used for this study. Columns En denote the English labels used, and columns Sp denote the Spanish labels used. While acknowledging that full accuracy across languages cannot be achieved, the team aimed to (i) validate

the order of the scales linguistically (e.g., "unlikely" refers to a lower probability than "somewhat likely"); (ii) use the simplest possible language in the labels to ensure respondent comprehension (see Issue 2 below); and (iii) preserve the original meaning in English of secure (+) or insecure (-) category.

However, achieving all three goals with the translation of the 5-point scale turned out to be difficult. In particular, the Spanish language equivalent for the category "Slightly likely" is either hard for respondents to understand or changes its meaning from being negative in terms of security of property rights in English to being positive in Spanish. Such changes have important implications for the data interpretation as discussed in the next section. This conclusion is in line with Liao et al. (2019) and Oberski et al. (2010), who argue that written and spoken language used in formulating the questions determine the reliability and validity of questions. They also argue that "cultural transportability of experimental and pretesting techniques cannot be assumed; it has to be tested" (Liao et al., 2019, p.119).

Considering the above difficulty with translation, an alternative set of labels for the 5-point scale was considered (see Table 8, rightmost column). However, the use of such an alternative turned out to be even more challenging as discussed under Issue 4 below.

**TABLE 8. LINGUISTIC COMPARISON OF SCALES MEASURING PTS**

| Prindex 4-point scale | | 5-point scale | | | | |
|---|---|---|---|---|---|---|
| **En** | **Sp** | **En** | **Sp** | **Sp: Literal translation** | **En: Reverse translation** | **Alternative by NORC** |
| 1. Very unlikely (+) | Muy poco probable | 1. Not at all likely (+) | Nada probable | *Nada probable* | Not at all likely (+) | 1. Not at all likely/Nada probable (+) |
| 2. Unlikely (+) | Poco probable | 2. Slightly likely (-) | Poco probable | *Ligeramente probable* | Unlikely (+) | 2. Somewhat unlikely/ Algo poco probable (+) |
| 3. Somewhat likely (-) | Algo probable | 3. Moderately likely (-) | Más o menos probable | *Moderadamente probable* | More or less likely (-) | 3. Neutral/Neutral (+/-) |
| 4. Very likely (-) | Muy probable | 4. Very likely (-) | Muy probable | *Muy probable* | Very likely (-) | 4. Somewhat likely/Algo probable (-) |
| | | 5. Extremely likely (-) | Totalmente probable | *Extremadamente probable* | Totally likely (-) | 5. Extremely likely/ Extremadamente probable (-) |

*Note: the highlighted results are discussed in the text. En = English, Sp =Spanish; assignment of +/- interpretation reflects authors' understandings gained from prior discussions at workshops and meetings.*

2. ***Levels of complexity in language***. The quality of communication depends on the complexity level of the instructions and the respective individual processing capacity level for such instructions. According to linguistic theories, the more educated a person is, the more likely it is that she or he is familiar with a wider and more complex vocabulary (number, use of grammatical rules, synonyms, etc.) and more formal language. The same is true about development of linguistic skills with a person's age. Although other levels can be considered, the following levels of complexity in language are commonly accepted in linguistics (Juboori & Muhaisin, 2017):

   - Firsthand, substandard, or popular level: sometimes used by individuals of lower levels of education;
   - Secondhand, everyday, colloquial level: used in specific environments or contexts such as a sport; and

- Super-standard level: not common for many speakers even if it is their mother tongue. It can be cultured (the result of education and age), technical (the result of specialization), and scientific.

Hence, for achieving optimal communication it is important to seek consistency between the level of language complexity and the comprehension levels of people with whom the specific communication is taking place.

The sample used for this study includes respondents with all levels of education and all age groups. For this reason, the team avoided using a literal translation of the 5-point standard unbalanced scale, as words like "slightly," "moderately," and "extremely" would increase the complexity in the language used and may lead to inconsistent interpretation of the labels by survey respondents (see Table 9, column Sp: Literal Translation).[24] The complexity of these words is reflected by the fact that they are adverbs, which refer to the way things are done and require a higher sophistication in communication. The initial feedback on the proposed set of labels was that the language was indeed too "academic" and is "not how the local people speak." Thus, the labels in the Sp column were used instead.

The same logic applies when we compare the clarity and complexity of symmetrical scales (with an equal number of positive and negative categories) vs. asymmetrical scales. The higher the disproportion between the positive and negative categories, the more complex language must be to describe the scale. For example, the original 5-point scale in English is understood (see footnote 16) to have one positive and four negative categories. This scale is more complex than the translated version with two positive and three negative categories (see the highlighted transformations in Table 8).

Issues 1 and 2 are interlinked and imply that the more complex the linguistic construct is, the more difficult it is to translate to a different language without losing the nuances of the original meaning.

3. ***Longer scales require higher language complexity***. An important message that stands out in Table 9 is that the 4-point scale is composed of labels with lower language complexity. The reason for this is that longer scales need to use more sophisticated and complex language to detect nuances between the categories. Similarly, the asymmetrical scale uses a larger number of negative categories, which requires use of more complex linguistic constructions to describe the nuanced differences among the categories.

**TABLE 9. LEVELS OF LINGUISTIC COMPLEXITY OF SCALES MEASURING PTS**

| Prindex 4-point scale | | | 5-point scale | | | | |
|---|---|---|---|---|---|---|---|
| En | Sp | Sp: Complexity | En | SP: Literal translation | Sp: Complexity | Sp | Sp: Complexity |
| Very unlikely | Muy poco probable | 1 | Not at all likely | *Nada probable* | 2 | Nada probable | 2 |
| Unlikely | Poco probable | 2 | Slightly likely | *Ligeramente probable* | 3 | Poco probable | 2 |
| Somewhat likely | Algo probable | 1 | Moderately likely | *Moderadamente probable* | 3 | Más o menos probable | 1 |
| Very likely | Muy probable | 1 | Very likely | *Muy probable* | 1 | Muy probable | 1 |

[24] The words chosen by the project team match up with the language used in questions from typical national surveys administered in Colombia by the national statistics office (DANE). The five terms "nada," "poco," "más o menos," "muy," and "totalmente" are used on the National Survey on Security and Coexistence (*Encuesta Ciudadana de Seguridad y Convivencia)* 2019. Similarly, the terms "nada," "muy," and "totalmente" are also used on the household survey *Gran Encuesta Integrada de Hogares* (2020) and the National Quality of Life Survey 2018 (*Encuesta Nacional de Calidad de Vida*). The questionnaires used by DANE can be accessed at https://www.dane.gov.co/files/anda/index.html.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | Extremely likely | *Extremadamente probable* | 3 | Totalmente probable | 1 or 2 |

*Note: The levels of linguistic complexity are classified as follows: 1) firsthand, substandard, or popular; 2) secondhand, everyday, colloquial; 3) super-standard. En for English, Sp for Spanish*

4. ***Ambiguity of neutral options when assessing likelihood.*** A symmetrical 5-point scale with a neutral middle point was considered as an alternative strategy to simplify the 5-point scale (Table 8, the rightmost column). The concept of neutrality, however, is problematic in surveys assessing the concept of likelihood of losing property rights. First, it requires a high complexity in language (3rd level or above). Second, it is essential to train enumerators in such a way that they can fully understand each option and have a clear reference of the circumstances in which each option is valid. The discussion with enumerators and survey designers revealed that in the case of probabilities, the notion of neutrality seems rather abstract, as it is difficult to have clear examples of the circumstances under which an event could be both likely or unlikely; or neither likely nor unlikely.[25] The notion of neutrality with respect to probability is, however, different from the neutrality with respect to PTS where people can describe their feelings as neither secure nor insecure. However, such application would require a different design of the survey question. As a result, this alternative was not considered beyond some initial discussions with the survey vendor and a small number of pre-tests. The evidence to support this conclusion was provided by the results of surveys conducted in the same municipalities during the same time period by a different project. The baseline data collection for an impact evaluation of LFP, being carried out by NORC, used the 5-point symmetrical scale to survey the PTS of landowners (summary statistics and comparison with Prindex data is presented in Annex 3). It shows that the "neutral" middle point was selected by a disproportionally low number of respondents. Krosnick (1991) cited by Krosnick & Presser (2010) also states that the middle point is more vulnerable to a satisficing effect – meaning that the probability of selecting a middle point is higher from people who intend to confirm an expected response rather than reveal their true attitude about the concept of interest. This effect is stronger among respondents with low cognitive skills and low motivation to complete the survey. This satisficing effect is not desirable, which supports the earlier conclusion.

5. ***Complexity of the notion of likelihood***. Future studies are encouraged to test alternatives to the notion of likelihood, as this term requires higher complexity of language in the context of measuring PTS, grounded in the existing broader body of literature in this area in which survey methodologists have grappled with this issue. Possible alternatives may include such terms as "possible" or "chance" with the respective changes across the scale categories and question formulation. Respective translation into other languages and mapping into secure/insecure categories should accompany such analysis. The linguistic complexity of the "likelihood" language was a point of debate at all stages of Prindex methodology development and a better solution has not yet been found.[26]

## 5.4 COMPARISON OF BINARY MEASURES OF PTS CONSTRUCTED FROM ALTERNATIVE 4-POINT AND 5-POINT ORIGINAL SCALES

As the two alternative measurement scales studied use a different number of categories for measuring PTS and are collected from two different sub-samples of respondents, comparing the PTS estimates across these scales requires rescaling to a common set of categories. For this comparison we selected a

[25] Examples under which losing rights is very likely (e.g., living in a municipality highly exposed to violent conflict, having disputes with neighbors, being in process of a divorce) or unlikely (e.g., documents are in place, there is political stability, harmony within the family) are clear yet the situations illustrating a neutral probability remain unclear.

[26] It should be highlighted, however, that the above discussion is linked directly to the use of the "likelihood" concept in the question. It may be the case that other concepts may be described reasonably well with the longer scales and include a neutral middle point. Also, scales that do not include labels for all the categories (e.g., end points only) can rely on simpler language and are likely to be less vulnerable to misinterpretation and translation inconsistencies.

binary measure of PTS (secure/insecure), as such a measure is commonly used in policy analysis and programming research, and is the simplest way to present PTS. There is, however, no single way to do this. Table 10 demonstrates several alternatives for constructing the binary measure of PTS, the corresponding values of PTS, and the test results for statistical differences between the two sub-samples for each of the proposed binary measures.

Two measures focused on security of PTS are considered (labeled S1 and S2 herein):

- **S1** is the preferred measure and considers tenure as secure if a respondent has selected any of the positive categories as described in the linguistic analysis section. On the 4-point scale, it corresponds to "Very unlikely" or "Unlikely" categories. On the 5-point scale, "Not at all likely" and "Slightly likely" are treated as secure because in Spanish the label for the second category has a positive meaning. The estimates show that respectively 64.9 percent and 64.7 percent of the population feel secure about their rights for the primary housing property and land attached to this property. The difference between the two alternatives is 0.2 percentage points and is statistically insignificant at a 10 percent significance level.
- **S2** is a measure of extreme security of property rights and includes only the top positive category in each scale. The difference between the two estimates is 1.8 percentage points and is also insignificant.

There are also at least three alternative ways to construct the measures of tenure insecurity (labeled I1, I2 and I3 herein):

- **I1** is the preferred measure of tenure insecurity and considers tenure as insecure if the respondents choose either of the two negative categories in the 4-point scale – "Somewhat likely" and "Very likely" – or any of the three categories in the 5-point scale (as the 5-point scale is asymmetrical) – "Moderately likely," "Very likely," and "Extremely likely." The respective estimates are 33.8 percent and 32.0 percent of the population who feel insecure about their property rights. The difference between the two estimates is statistically insignificant at a 10 percent significance level.
- Measures of insecurity **I2** and **I3** are designed to capture extreme insecurity and demonstrate large statistically significant differences in the level of tenure insecurity between the two sources of data, which potentially could lead to very different policy implications.
    - **I2** includes the top negative value of "Very likely" on the 4-point scale and the top two negative values of "Very likely" and "Extremely likely" on the 5-point scale.
    - **I3** includes the top negative value on both scales.
- Comparison of the measures from the current study with similar measures from the Prindex national data collected in 2018 (see Section 3 for discussion) highlights that the measures of insecurity are likely to be more affected by non-responses than the measures of tenure security (e.g., people who feel insecure are more likely to refuse answering the question).

In summary, the two sub-samples of data (collected with 4-point scale and with 5-point scale) demonstrate a similar level of tenure security, but the levels of tenure insecurity depend on the method used to construct these measures. The analysis may become even more complicated if a symmetrical 5-point scale with a neutral middle point is used instead (not used for the data collection in the current study though some comparison is presented in Annex 3). Such a symmetrical scale would require establishing a rule about treating the middle point as secure or insecure, or splitting 50/50 or in some other proportional manner between secure and insecure. As presented in Table 10, a large share of observations are located around the middle point of the tenure security distribution. That implies that

the rules for allocation of observations within the middle category between secure and insecure may significantly affect the estimates and their comparability with alternative measures of PTS and across countries, regions, and municipalities.

**TABLE 10. ALTERNATIVE MEASURES OF PTS COMPARED IN THIS STUDY**[27]

| Measure | 4-point scale | | 5-point scale | | Difference/ Significance |
|---|---|---|---|---|---|
| | Construction | Statistics, % | Construction | Statistics, % | mean(4pt) - mean(5pt) [Pr(diff=0), %] |
| Secure tenure (s1) | [(1) Very unlikely + (2) Unlikely)] / Total number of observations | 64.9 | [(1) Not at all likely+ (2) Slightly likely] / Total number of observations | 64.7 | 0.2 [92.3%] |
| Secure tenure (extreme) (s2) | (1) Very unlikely / Total number of observations | 33.8 | (1) Not at all likely / Total number of observations | 32.0 | 1.8 [34.7%] |
| Insecure tenure (i1) | [(3) Somewhat likely + (4) Very likely] / Total number of observations | 33.0 | [(3) Moderately likely + (4) Very likely + (5) Extremely likely] / Total number of observations | 33.9 | -0.9 [67.7%] |
| Insecure tenure (extreme) (i2) | (4) Very likely / Total number of observations | 10.2 | [(4) Very likely + (5) Extremely likely] / Total number of observations | 14.4 | -4.1*** [0.5%] |
| Insecure tenure (extreme) (i3) | (4) Very likely / Total number of observations | 10.2 | (5) Extremely likely / Total number of observations | 2.0 | -8.2*** [0.0%] |
| Non-response | (Don't know + Refused) / Total number of observations | 2.1 | (Don't know + Refused) / Total number of observations | 1.4 | 0.7 [15.8%] |

*Note: *** - Significant at 1%; **- Significant at 5%; * - Significant at 10%; sampling weights are applied*

## 5.5 COMPARISON OF NON-RESPONSE RATE ACROSS THE 4-POINT AND 5-POINT SCALES

In a similar way we compare the non-response rate between the two scales. The bottom row in Table 10 demonstrates that data collected with both 4-point and 5-point measurement scales are associated with statistically similar shares of non-responses (2.1 percent and 1.4 percent, respectively). The "Don't know" and "Refuse" are both treated as non-responses. For comparison, the "worry" question produces a lower non-response rate, about 0.3 percent (see Table 7)

## 5.6 SUMMARY OF COMPARATIVE ANALYSIS OF 4- AND 5-POINT MEASUREMENT SCALES

In summary, the descriptive analysis of data collected with 4- and 5-point scales demonstrate the following:

- Binary measures of PTS (secure/insecure) based on both sub-samples of data demonstrate a **similar** level of tenure security.
- Binary measures of PTS of tenure insecurity demonstrate significantly different results depending on how this measure is constructed.

[27] As noted previously, we have added Annex 3 to take a comparative look with the findings of the LFP impact evaluation baseline data (implemented by NORC).

- Data collected with both scales demonstrate a **similar** unit non-response rate.
- The language used to describe the labels for two alternative scales demonstrates a **different** level of complexity, which may lead to a higher degree of misinterpretation of responses when the current version of the 5-point scale is used.
- The 5-point scale is **more vulnerable** to misclassification when converted into the binary measure of PTS (based on both symmetrical and asymmetrical scales) as the middle category(ies) may be incorrectly classified as secure or insecure and this classification error may be different in different languages.

The above list highlights some differences in results from descriptive analysis of data collected using the two different measurement scales (the two sub-samples of data). However, descriptive statistical analysis does not shed light on data quality differences that may result from the differences in the measurement scales (e.g., how well each scale captures the true variation in the level of tenure security across population).

# 6.0 COMPARATIVE ASSESSMENT OF DATA QUALITY (RELIABILITY AND VALIDITY) AND BIAS IN USE OF THE TWO ALTERNATIVE PTS MEASUREMENT SCALES

Following on the descriptive analysis presented in Section 5, in this section we present a more rigorous analysis of which scale produces a higher quality of data for PTS.

Both the 4-point and 5-point scales are attempting to capture a respondent's real, unobservable level of tenure security, the degree to which they truly feel secure or insecure about their property rights. This is referred to as their "true perception of tenure security" (TPTS). Yet numerous studies suggest that there is often a disconnect between a respondent's internal assessment of their level of tenure security and how they report their perception of security on a given measurement scale. Thus, variation in responses between two respondents could reflect variations in the underlying TPTS, systematic measurement error associated with a particular survey instrument, or random error. In order to assess the reliability and validity of the two scales, the study uses a regression model, based on alternative measures of PTS asked in the survey, to develop a proxy measure, or predicted value, of the underlying TPTS. This proxy is estimated based on 11 alternative survey measures of PTS. The correlation between the predicted value of TPTS and the PTS measures on the two alternative scales is used as an estimate of data quality for each instrument. Researchers also look at the correlation between respondent's answer to the 4- and 5-point likelihood questions, and another commonly used PTS question: "*How worried are you that you could lose the right to use this property, or part of this property, against your will in the next 5 years?*" This "worry" question is used as another proxy measure for how well the alternative scales are capturing the underlying value of TPTS. The study then looks at the distribution of the predicted values of TPTS to assess which number of categories best fits the natural buckets in the data. Then the predicted category of the TPTS value is compared with the category given by the respondent to the survey question to assess which scale produces the highest proportion of correctly classified observations. This is used as an estimate of instrument bias.

The following sections outline the theoretical underpinnings of the analytical approach taken and highlight the key findings of each subsequent test of data quality and bias.

## 6.1 DATA QUALITY ASSESSMENT METHODOLOGY

The literature on survey design provides some mixed and indirect evidence with respect to the data quality of 5-point and 4-point measurement scales. Revilla et al. (2014) suggest that a 5-point scale produces data of a higher quality in comparison to longer (7- or 11-point) scales. The Organization for Economic Co-operation and Development (OECD, 2013) found that binary scales are more robust to cultural differences in the response patterns. Krosnick and Presser (2010) in their review point out that 4-, 5-, and 6-point scales demonstrate the highest reliability versus shorter scales and no gain in reliability compared to longer scales. However, Alwin (2010) mentions that the performance of a particular element of design is likely to be specific to a particular context, which highlights the importance of studies on survey question design specific to PTS and to country/institutional environments. Moreover, the combination of different elements of design (number of points on scale, presence of a middle point, symmetrical or asymmetrical set of response options, presence of a "don't know" option, clarity of labels, etc.) may have a reinforcing effect on the data quality. Thus, this study

compares the complete design of the two alternative measurement scales, presented in Boxes 1 and 2 (Section 2, above), rather than only the effect of specific elements of these.

The literature on survey designs suggests that the data quality ($q_{ij}$) of a measure (*i),* measured with an instrument (*j)* could be presented as:

$q^2=(r*v)^2$

where (r) represents reliability and (v) is a construct validity of the instrument (Alwin, 2007; Alwin & Krosnick, 1991). The square of the data quality indicator ($q^2_{ij}$) represents the amount of variance in a measure of PTS, which is explained by the latent trait of interest (Liao et al., 2019), which in case of this study is the true perception of tenure security (TPTS).

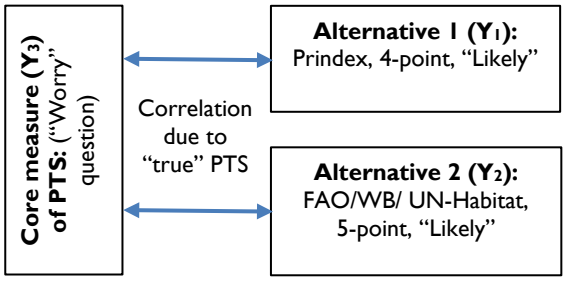
Variation in any measure $Y_{ij}$ (in our case, based on Prindex or FAO/WB/UN-HABITAT instruments) reflects the variation in the underlying concept of TPTS, systematic measurement error (associated with application of a particular survey instrument $M_j$), and a random error ($e_{ij}$).

$Yij = r_i*v_jTPTS + r_i*m_j*M_j+e_{ij}$

The reliability ($r_i$) corresponds to the variance of the measure $Y_{ij}$ explained by the underlying variable of interest - TPTS (Reliability is $r^2 = 1 - (Var(e)/Var(Y))$. The construct (internal) validity ($v_j$) tells us what part of this variation is valid (caused by the TPTS), and invalid (caused by the application of a particular instrument $m_j=1-v_j$).

Estimation of the above parameters is not feasible directly, as the TPTS is not observable. Instead, a multitrait-multimethod (MTMM) methodology (e.g., Saris et al., 2004, Alwin, 2010) is used in the literature. MTMM provides a way to estimate the data quality parameters if the data on the same concept are collected from the same respondents with at least three alternative instruments (methods). In our study, the alternative instruments are the following (see Figure 3):

**FIGURE 3. SPLIT-BALLOT MULTITRAIT-MULTIMETHOD METHODOLOGY**




- **Alternative 1** ($Y_1$): the instrument is based on the symmetrical Prindex 4-point scale: "Very unlikely [positive PTS], unlikely [positive PTS], somewhat likely [negative PTS], very likely [negative PTS];"
- **Alternative 2** ($Y_2$): the instrument is based on the asymmetrical FAO/WB/UN-HABITAT 5-point scale: "Not at all likely [positive PTS]; slightly likely [negative PTS]; moderately likely [negative PTS]; very likely [negative PTS]; extremely likely [negative PTS];" and
- **Core measure** ($Y_3$): an alternative instrument measures PTS based on the "worry" question related to losing property rights (used in Prindex studies, see Box 3).

**Box 3. Prindex, 2018 questionnaire**

Q45. How worried are you that you could lose the right to use this property, or part of this property, against your will in the next 5 years?

1. Not worried at all
2. Not worried
3. Somewhat worried
4. Very worried
5. (Don't know)
6. (Refused)

In all the alternative measures, the central questions refer to the tenure security of the main housing property and the land attached to it. To test the difference in the scale designs, Alternative 1 and Alternative 2 use the same likelihood-based

question from the Prindex methodology (presented in Box 1), but on two different scales as discussed above.

The essence of this method is that the correlation between the alternative measures would include the common factor caused by the variance in the unobservable TPTS and distortions caused by the systematic measurement error associated with a specific instrument as presented below:

$$Cor(Y_1\ Y_2)= r_1v_1Var(TPTS)v_2r_2+r_1m_1m_2r_2$$

All these parameters can be estimated using a structural equation model (Saris & Gallhofer, 2007, 2014) under a set of assumptions about a lack of correlation among the error terms and the method effects.

There are, however, several practical problems with implementing this method in the assessment of PTS, which require modification of the standard MTMM approach:

1. Within the budget of this study, it was not practical to collect data on all three measures within the same interview (particularly the very similar questions under Alternatives 1 and 2). The budget provided for 20 – 25 minute interviews, which is not enough time to avoid contaminating one question with a previous version of the same question (e.g., Y1 vs. Y2). Instead, this study randomly split the representative samples of each municipality population into two sub-samples and administered the two alternative instruments (for constructing Y1 and Y2) in the respective sub-samples. Thus, each respondent would have to answer only two questions: Y3, and either Y1 or Y2. Such an approach represents a variation of the split-ballot MTMM (SB-MTMM) (Saris et al., 2004; Revilla et al., 2014; Revilla & Saris, 2013).

2. For the same reason, this study was limited in applying the alternative instruments to at least three alternative concepts (traits), which is necessary for estimating the instrument effects ($m_j$). Instead, we used additional measures of PTS available in the Prindex survey (see the list and respective descriptive statistics in Table 6, Section 5) to estimate a predicted value of PTS for each respondent as a common factor among all the alternative measures of PTS using the Generalized Structural Equation Model (GSEM) framework for predicting a latent variable. This predicted value is used as a proxy for TPTS. The correlation between the predicted value of TPTS and the measures of interest (Y1 or Y2) are used as estimates of data quality associated with each instrument.

3. The systematic correlation among the measures of PTS is likely to be caused not only by the TPTS, but also because of some common design features. In particular, several measures of PTS rely on different versions of the "worry" based question. Also, several measures are using different versions of the 4-point scale (see Table 11 for details). We estimate the effect of these common design parameters in addition to the common TPTS effect (see questions that feed into the GSEM estimation model in Table 11 and results in Table 12) and use the TPTS prediction net of the above design effects in further analysis.

**TABLE 11. DESCRIPTION OF THE STRUCTURAL MODEL FOR TPTS**

| Measure of PTS | Label | TPTS | Worry | 4pt |
|---|---|---|---|---|
| ($\mathbf{Y_1}$) In the next 5 years, how likely or unlikely is it that you could lose the right to use this property, or part of this property, against your will? (4-point) | **Likely Question – 4 Pt** | + | | + |
| ($\mathbf{Y_2}$) In the next 5 years, how likely or unlikely is it that you could lose the right to use this property, or part of this property, against your will? (5-point) | **Likely Question - 5 Pt** | + | | |
| ($\mathbf{Y_3}$) How *worried* are you that you could lose the right to use this property | **Worry Question** | + | + | + |

| | | | | |
|---|---|---|---|---|
| If you had a *disagreement/dispute with your family*, how worried would you be that your right to stay in this property would be taken away from you | Family Dispute | + | + | + |
| If *someone else in your family lost their job*, how worried would you be that your right to stay in this property would be taken away from you | Family Job Loss | + | + | + |
| If a *company tried to take over the land* your dwelling is on against your will, how worried would you be that your right to stay in this property would be taken away from you | Company Taking | + | + | + |
| If *another person or group claimed ownership*, how worried would you be that your right to stay in this property would be taken away from you | Competing Claim | + | + | + |
| If *somebody else fraudulently sells the property*, how worried would you be that your right to stay in this property would be taken away from you | Fraud | + | + | + |
| If a *neighbor initiates a boundary dispute*, how worried would you be that your right to stay in this property would be taken away from you | Boundary Dispute | + | + | + |
| If a *disagreement arose with local/customary authorities* (e.g., officials/chiefs, elder), how worried would you be that your right to stay in this property would be taken away from you | Local Authority Dispute | + | + | + |
| In general, how well do you think people in this country are protected when it comes to their property rights? | General Protection | + | | + |

The experimental design used for data collection controls effectively for the rest of the possible sources of data quality differences between the 4- and 5-point sub-samples except of the difference in PTS measurement scales, as explained above in Section 4.

## 6.2 RESULTS FROM ESTIMATION OF THE STRUCTURAL MODEL TO PREDICT TPTS

Estimation results for the structural model described above (Section 6.1) are represented by the path coefficients between the latent variables and the respective survey-based measures of PTS, as well as the associated estimates of variation. Table 12 presents the partial results for the measure of latent TPTS. The complete set of estimation results is presented in Table A1.1 in Annex 1.

**TABLE 12. PARTIAL ESTIMATION RESULTS FOR PATH COEFFICIENTS TO THE "TRUE" PTS**

| Variable | TPTS | |
|---|---|---|
| | Path Coefficient | Odds Ratio |
| Likely question – 4-point | 2.12 | 8.36 |
| Likely question – 5-point | 2.00 | 7.42 |
| Worry question | 1 | 2.72 |
| Family dispute | 0.57*** | 1.77*** |
| Family job loss | 0.58*** | 1.78*** |
| Company taking | 0.73*** | 2.07*** |
| Competing claim | 0.93*** | 2.53*** |
| Fraud | 0.75*** | 2.12*** |
| Local authority dispute | 0.63*** | 1.88*** |
| Boundary dispute | 0.59*** | 1.80*** |
| General protection | -0.14*** | 0.87*** |

*Note: sampling weights are applied; *** significant at 1% significance level*

It should be noted that the latent variables (such as the TPTS) do not have any associated measurement scales because they are unobservable. Thus, their values must be normalized. The path coefficient to one of the observed variables is set equal to one for this purpose. This type of variable is called an anchor variable. This variable sets the direction for variation of the corresponding latent variable. The paths to other variables are measured relative to the anchor variable. As all the observed variables are categorical, the path coefficients are estimated with the ordered logit estimator. Table 12 presents both the path coefficients and their exponentiated values, which can be interpreted as odds ratios (OR). An OR greater than one implies that a unit increase in the alternative measure of PTS increases the chance of observing a higher value of TPTS, while a value below one decreases this chance.

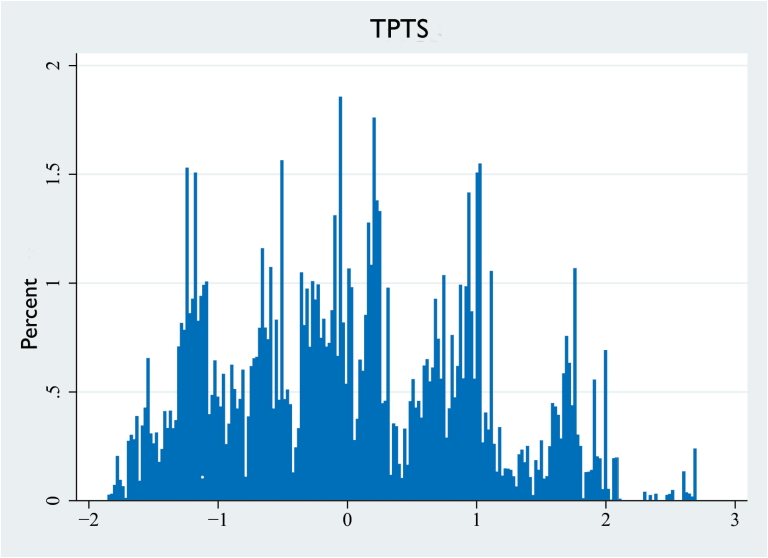
For the latent TPTS, the "worry" question ($Y_3$) is set as an anchor. The statistically insignificant values of the path coefficients and the ORs for both likely questions (Likely 4-pt and Likely 5-pt) indicate that the values in the "worry" question and both likely questions vary along the same vector. All other measures of PTS demonstrate significant ORs with TPTS, which indicates that these variables contribute to prediction of TPTS beyond the contribution of the "worry" anchor variable. The magnitude of the OR reflects the relative contribution of each variable to the common variation in TPTS along the direction of the anchor variable.

## 6.3 PREDICTION OF TPTS

As the TPTS is unobservable, some additional assumptions are needed. It is assumed that the predicted value of TPTS is averaged at zero. The distribution of the predicted values of TPTS is presented in Figure 4. The TPTS is constructed as a continuous variable and uses values in a range from -1.86 to 2.7 (Table 13), where more negative values are associated with a higher level of perceived security and more positive values correspond to higher insecurity. As demonstrated in Table 13, the distribution of TPTS is similar across the two sub-samples as assessed by the difference in means.

As was argued above, the predicted values of TPTS may be affected by the common design features among the measures of PTS used for the prediction: "worry"-based questions and 4-point scales. Table A1.2 demonstrates that these two design effects jointly explain about 3.6 percent of variation in the predicted TPTS. Residuals from the presented linear regression represent the predicted TPTS net of the common design effect (measurement error) and are used in further analysis.

**FIGURE 4. DISTRIBUTION OF PREDICTED VALUES OF TRUE PTS**




*Note: sampling weights are applied*

**TABLE 13. SUMMARY STATISTICS FOR THE PREDICTED VALUES OF TPTS**

| Sample | Variable | Obs | Mean | Std. err. | Min | Max |
|---|---|---|---|---|---|---|
| 4-pt | TPTS | 2,275 | 0.02 | 0.03 | -1.86 | 2.10 |
| 5-pt | TPTS | 1,759 | 0.03 | 0.03 | -1.82 | 2.70 |

*Note: sampling weights are applied*

## 6.4 DATA QUALITY (RELIABILITY AND VALIDITY)

As mentioned above, the data quality of a particular measure of PTS can be presented as a percentage of variation in this measure explained by the TPTS. We use the square of the Pearson correlation coefficient as a measure of data quality (Coenders & Saris, 1995). Table 14 presents the results, which shows the following:

- The data quality of both measures of PTS – collected using the 4- and 5-point scales – produce a comparable level of data quality. The TPTS explains respectively 89 percent and 93 percent of variation in the "likely" based measures of PTS (see numbers in bold in Table 14). The larger value of data quality for the 5-point scale demonstrates that this scale is able to capture some additional variation in the TPTS in comparison with the 4-point counterpart.
- Both measures of PTS based on the "likely" question perform much better than the alternative proxy measures of PTS (worry, etc.) and should be treated as a preferred measure of PTS among the alternatives available in Prindex data set.
- The measure of PTS based on "worry" question ($Y_3$) is the second best. The TPTS contributes only about 28 percent to the variation in this measure. This relatively low reliability of the "worry" based measure of PTS explains a relatively low correlation between the "likely" and "worry" based measures of PTS.
- The negative correlation with some of the measures represents that the scales use different directions.

**TABLE 14. ESTIMATED DATA QUALITY OF PTS MEASURES AS MEASURED BY CORRELATION WITH TPTS**

| Sample: | 4-pt | | 5-pt | |
|---|---|---|---|---|
| Measures of PTS based on: | Corr with TPTS | Data quality | Corr with TPTS | Data quality |
| Likely question | 0.94 | **0.89** | 0.97 | **0.93** |
| Worry question | 0.53 | 0.28 | 0.53 | 0.28 |
| Competing claim | 0.26 | 0.07 | 0.29 | 0.08 |
| Fraud | 0.23 | 0.05 | 0.24 | 0.06 |
| Boundary dispute | 0.19 | 0.03 | 0.19 | 0.04 |
| Family dispute | 0.12 | 0.02 | 0.14 | 0.02 |
| Local authority dispute | 0.16 | 0.03 | 0.16 | 0.03 |
| Family job loss | 0.07 | 0.01 | 0.16 | 0.03 |
| Company taking | 0.19 | 0.04 | 0.23 | 0.05 |
| Government taking | 0.18 | 0.03 | 0.17 | 0.03 |
| General protection | -0.13 | 0.02 | -0.16 | 0.02 |
| Had a dispute | -0.08 | 0.01 | -0.14 | 0.02 |
| Was evicted | -0.08 | 0.01 | -0.10 | 0.01 |

*Note: sampling weights are applied*

## 6.5 ASSESSMENT OF BIAS

Conceptually, bias (B) in a measure of PTS is associated with a systematic over- or underprediction of the true value of PTS. It can be presented as:

$B=E(PTS_j)-E(TPTS)$

where *j* is a survey measure of PTS (based on 4-point or 5-point scales) and E(TPTS) is the expected value (E) of a latent PTS measure (equal to zero by construction). However, in the case of a categorical variable, such as the "likely"-based survey measures of PTS, the bias is estimated as a percentage of incorrectly classified observations (Kantardzic, 2019). For such estimation, the predicted continuous value of TPTS (see section 6.1) has to be classified respectively into four or five categories and then compared to the values recorded with the survey instrument. We can also assess if converting the original data on PTS into a binary measure (secure/insecure) would increase or decrease the bias. In parallel, we assess what is the most informative number of categories that can be applied for classification of TPTS. It is assumed that a higher quality of data can be achieved if the measurement scale has the same number of categories as the TPTS and the labels help to allocate observations into the same classes as in the distribution of TPTS.

### 6.5.1 CLASSIFICATION OF TPTS

For classification of TPTS we apply latent category analysis methodology. For a selected number of categories of perceived tenure security, this method predicts the probability that an observation belongs to each category based on the value of TPTS and then selects the category that is associated with the highest probability. The quality of resulting classification can be assessed with the log-likelihood measure (LL) or one of the information criterion (IC) – Bayesian information criterion (BIC) or Akaike information criterion (AIC).[28] It should be noted that there is no perfect allocation of observations into the classes and some uncertainty in allocation of some observations remains as long as the number of classes is lower than the number of observations.

We estimate the classification into 2, 3, 4, 5, 6, 7, and 8 categories and the corresponding values of the classification quality measures. Results are presented in Figure 5. The lines in Figure 5 represent the values of the respective optimization criteria for each classification (the lines for BIC and AIC overlap). The results imply that while there are multiple ways to classify the TPTS, grouping observation into five categories is associated with the best values of the optimization criteria, which explains the higher correlation of TPTS with the 5-point measure presented in Table 14. The lowest value is associated with the binary classification.

We should note, however, that the presented results describe the characteristics of a specific dataset and are not robust if we repeat the classification of TPTS for other sub-populations (e.g., municipalities, rural or urban area, etc.). To demonstrate this point, Figure 6 presents the classification results for each of the four municipalities. This result implies that the optimal number of categories is different in each municipality. It is five categories in Puerto Lopez and Tolu Viejo municipalities, six categories in Zaragoza, and four categories in San Juan Nepomuceno. The results for the total sample present the weighted average of the sub-population results.

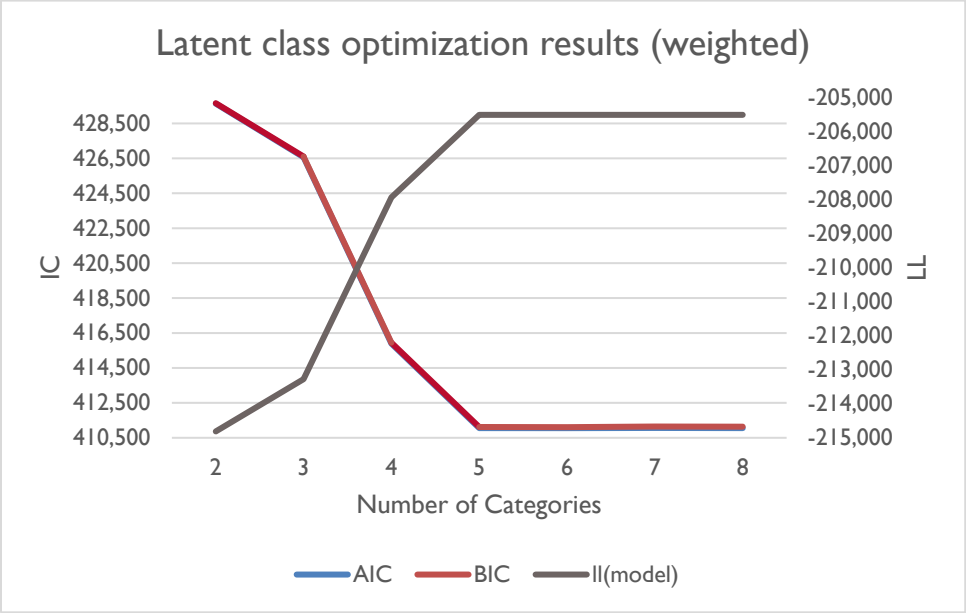
Annex 2 presents the classification results with respect to other characteristics. It shows that there is no difference in classification results with respect to gender (Figure A2.2) or income sufficiency levels

[28] The BIC and AIC are statistics for model optimization. The lower the AIC or BIC, the better is the predictive power of the model.

(Figure A2.4). The classification with respect to form of tenure shows that the optimal number of categories for owners is six, for renters it is eight, and for other forms of tenure it is five (Figure A2.3).

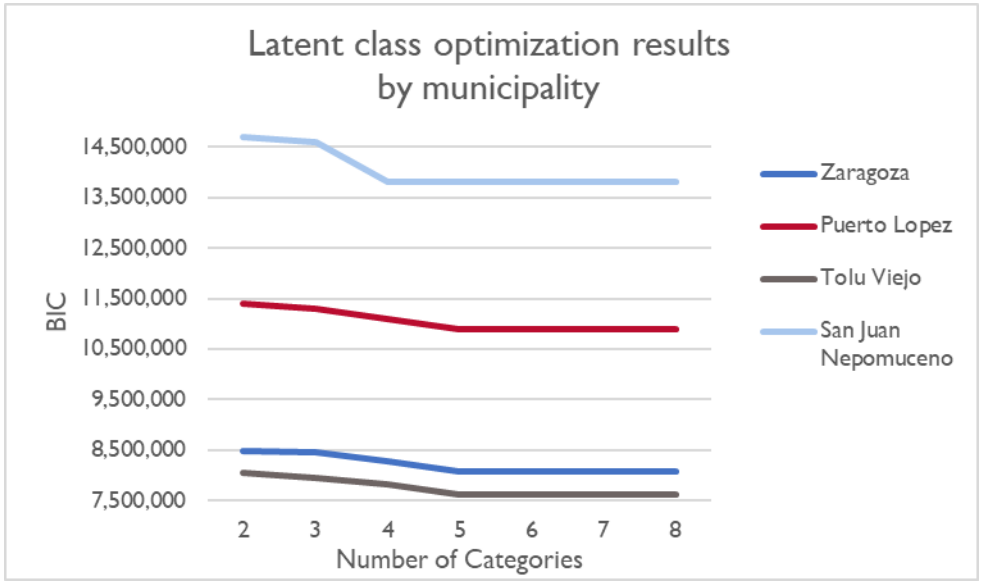
In summary, there is no one right or best way to classify the observations on TPTS into four or five categories, and the performance of a particular measure of PTS must be compared against the respective classification in the TPTS with the same number of categories: four, five, or two. However, such comparison is not feasible prior to data collection and the choice may be driven by practicalities and comparability with other sources of data.

**FIGURE 5. LATENT CLASS ANALYSIS FOR TOTAL SAMPLE**




*Note: sampling weights are applied. LL = log likelihood; IC = information criterion; BIC = Bayesian information criterion; AIC=Akaike information criterion.*

**FIGURE 6. LATENT CLASS ANALYSIS BY MUNICIPALITY**



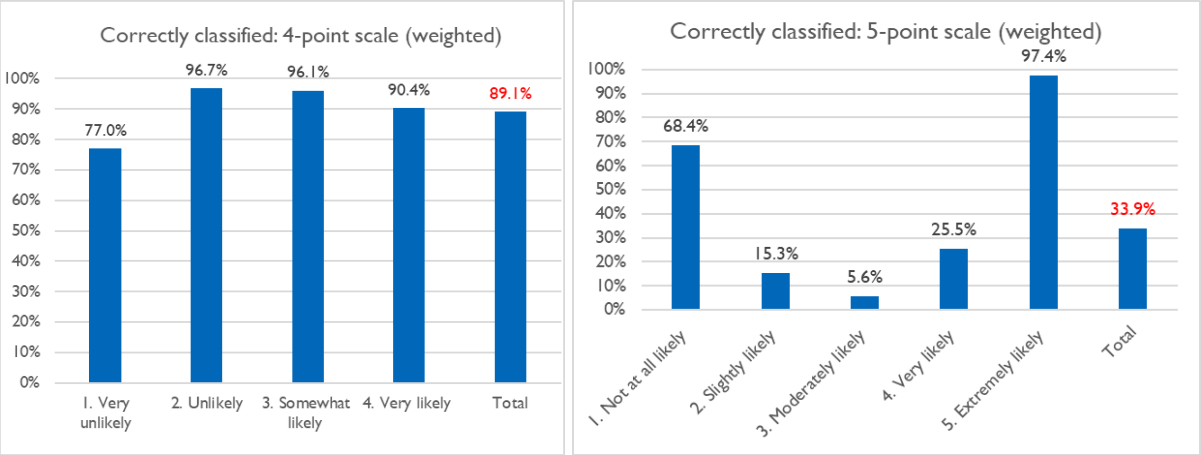

*Note: sampling weights are applied. BIC = Bayesian information criterion*

### 6.5.2 MISCLASSIFICATION AS A MEASURE OF BIAS

The assessment of bias in measures of PTS is based on the percentage of matches in classification between the TPTS and the respective measures of PTS based on 4-point and 5-point scale data. Figure 7 presents the results. It shows that the PTS measure based on a 4-point symmetrical scale classifies 89.1 percent of observations into the same categories as the 4-category classification of TPTS. With respect to individual categories, the percentage of correct classification varies from 77.0 percent to 96.7 percent. It should be noted that classification of the extreme categories performs less well than the middle categories. This observation is in line with recommendations from the linguistic analysis, which are presented in Table 15.

Regarding the performance of 5-point asymmetrical scale, only 33.9 percent of observations are classified into the same categories as in the optimal 5-category classification of TPTS. The middle categories perform particularly poorly with matches between 5.6 percent and 25.5 percent of observations. We can see that the observations recorded in the middle categories of the 5-point scale systematically have a higher level of insecurity in comparison to the comparable categories in the TPTS classification (the level of tenure security is overpredicted by the 5-point scale for the middle part of the distribution). The weaker link between the classification with the 5-point scale and TPTS in comparison with the 4-point scale classification is confirmed by the Cramér's V statistics, which is lower for the cross-tabulation of 5-point scale data (0.77 vs. 0.89 for the 4-point scale) This means that the asymmetrical nature of 5-point scale and complexity of language are likely to contribute to mismatch between the true feeling of tenure security level and the proposed categories (labels).[29] This suggests that the labels on the 5-point scale would benefit from a redesign.[30]

**FIGURE 7. SHARE OF CORRECTLY CLASSIFIED OBSERVATIONS BETWEEN THE SURVEY DATA AND CLASSIFICATION OF TPTS**




[29] While the above results demonstrate a relatively poor performance of the 5-point asymmetrical scale, our analysis is incomplete if we do not recognize the potential benefits of the 5-point scale. First, it can potentially have a higher reliability if the TPTS can be optimally classified into more than 4 categories (suppressing the true variation into the smaller number of categories would lead to a loss of information). However, if the optimal number of categories is less than 5, the longer scale would generate a measurement error. Second, if the true distribution of PTS is not symmetrical, a longer (and asymmetrical) scale can potentially better capture such distribution, which is likely to be a reason behind the design of FAO/WB/UN-Habitat scale.

[30] Even with a redesign, any 5-point scale would be asymmetrical by definition unless the middle point was neutral, which adds to the complexity of interpretation as presented in the linguistic analysis above. Redesign may include the scale with a larger number of positive categories than negative or a scale where only the end points are labelled.

**TABLE 15. POSSIBLE ADJUSTMENTS FOR SPANISH LABELS FOR PTS 4-POINT SCALE**

| Prindex 4-point scale | | |
|---|---|---|
| **English** | **Spanish: current** | **Spanish: alternative** |
| Very unlikely | Muy poco probable | Nada probable |
| Unlikely | Poco probable | Poco probable |
| Somewhat likely | Algo probable | Algo probable |
| Very likely | Muy probable | Totalmente probable |

Finally, we assess if the reclassification of the collected PTS data into binary variables (secure/insecure) helps to reduce the bias. Table 16 presents the results. It demonstrates that the measure of tenure security based on a 4-point scale data overestimates the level of security of property rights by 2.4 percentage points in comparison to the binary classification of TPTS, while the 5-point scale data underpredicts the level of security by 1.1 percentage points. Similarly, if we compare the measures of insecurity, the 4-point scale data underestimates this statistic by 4.5 points and 5-point scale data only by 0.3 points. Overall, the 4-point scale data provides for 94 percent match in binary classification between the survey data and the predictions based on TPTS. The 5-point scale data provides for 99 percent match.

**TABLE 16. QUALITY OF BINARY CLASSIFICATION OF PTS**

| Binary outcome | 4-pt | 5-pt |
|---|---|---|
| TPTS Secure (*) | 62.5% | 65.8% |
| TPTS Insecure (*) | 37.5% | 34.2% |
| Likely Secure (s1) | 64.9% | 64.7% |
| Likely Insecure (i1) | 33.0% | 33.9% |
| Bias Secure, ppt | 2.4% | -1.1% |
| Bias Insecure, ppt | -4.5% | -0.3% |
| Correctly classified binary | **94%** | **99%** |

Note: *(*) based on 2-category classification of TPTS; sampling weights are applied;* the highlighted results are discussed in the text

In summary, the re-classification of data collected with each alternative scale into a binary measure of PTS allows for a reduction in bias in estimates of tenure security level, as the misclassification among the smaller categories is eliminated. However, the quality of binary classification in the 5-point scale is conditional on the correct interpretation of the middle points in the distribution, which may differ from one language to another and is based on assumptions about the distribution.

## 6.6 ASSESSMENT OF DIFFERENCES IN QUALITY OF PTS MEASURES USING AN ONLINE TOOL FOR PREDICTION OF SURVEY QUALITY

The robustness of the findings of this report was also checked with the help of an alternative analytical model (Survey Quality Predictor 2.1 [SQP]) that evaluates the quality of survey instruments based on the meta-analysis of around 3,000 survey design experiments (Saris & Gallhofer, 2007; Saris et al., 2011). The SQP is available online at http://sqp.upf.edu/ and permits use of 46 generic characteristics of survey questions and the related scales to predict the data quality, reliability, and validity.

Comparison of the PTS measures based on 4- and 5-point scales data uses the English version of the questionnaire. The only differences in the considered characteristics are related to the following:

- Number of categories (four vs. five);

- Symmetry of measurement scale (symmetrical vs. asymmetrical); and
- Number of syllables in answer scale (20 vs. 26 as the 5-point asymmetrical scale requires more nuanced labels).

The results are presented in Table 17. It shows that the measure based on 5-point scale is associated with a lower validity and overall data quality in comparison to the 4-point counterpart. However, this difference is relatively small. Among the top recommendations for increasing data quality for the 5-point scale-based measure provided by the SQP model is to change the number of categories from five to four.

**TABLE 18. PREDICTED DATA QUALITY**

| Data quality measure | | 4-pt | 5-pt |
|---|---|---|---|
| Reliability = 1 - random error | $r^2$ | 0.703 | 0.709 |
| Validity = 1 - method effect | $v^2$ | 0.941 | 0.872 |
| Quality = reliability ($r^2$) x validity ($v^2$) | $q^2$ | 0.662 | 0.619 |

*Source: based on predictions from SQP2.1 available at http://sqp.upf.edu*

# 7.0 OTHER LESSONS LEARNED ABOUT MEASURING PTS

In addition to the experiment with the use of alternative scales for measuring PTS, the survey used in Colombia has several other innovations in assessment of tenure security, which are presented in this section. First, we discuss questions related to the right to benefit from common resources and public infrastructure. Second, we introduce questions assessing the enumerator's observations over the influence of other household members present during the interview.

## 7.1 INSECURITY IN TERMS OF ACCESS TO COMMON POOL RESOURCES AND PUBLIC INFRASTRUCTURE

The right to benefit from common pool resources (e.g., forests, water, pasture) and public infrastructure (e.g., roads, dams, public irrigations systems) is an important determinant of individual well-being. Similarly, access to public goods and common pool resources may be among the reasons why people decide to live in a certain region and they may rely on these resources for their livelihood. In this study we included new questions in the Prindex survey with the aim of ***assessing what share of the population benefits from access to such resources as well as the respondents' perceptions over the likelihood of losing their access.***

Most people in the surveyed municipalities (95 percent) have access to roads and approximately half have access to sources of water such as lakes, rivers, etc. (Table 18). However, 14 percent of the population perceives that it is likely that they could lose access to water in the next five years, which makes bodies of water the most insecure common resource. However, in terms of percentage of actual users, pastures are more insecure.

**TABLE 18. ACCESS TO COMMON RESOURCES AND PUBLIC GOODS AND LIKELIHOOD OF LOSING ACCESS**

| Common resources | Have access | Likelihood of losing access | |
|---|---|---|---|
| | % over total population | | % of users |
| Forests or natural parks | 35% | 8% | 23% |
| Sources of water | 55% | 14% | 25% |
| Public roads | 95% | 9% | 9% |
| Prairies/pastures | 15% | 4% | 27% |

## 7.2 INDIVIDUAL PERCEPTIONS CAN BE CAPTURED DESPITE POWER DYNAMICS WITHIN HOUSEHOLDS

One of the key features of the Prindex methodology is related to the random selection of respondents within a household, instead of interviewing the head of household, as is traditionally done in household surveys. This approach provides for the sample being representative of the adult population of each municipality. Also, it allows for the detection of different degrees of vulnerabilities across groups (i.e., young people, women, family members of the owners, etc.) and sources of tenure insecurity (including intra-household differences).

Nevertheless, qualitative observations from field accompaniment trips in previous studies identified difficulties in capturing *individual* perceptions, especially in the most vulnerable groups. Situations such as the parents overhearing the interview of a randomly selected son or the head of household (male) accompanying his wife during the interview raised questions on the extent to which the surveys were

capturing individual perceptions. The Prindex team decided to assess the influence of presence of other individuals in survey responses.

At the end of each completed interview, enumerators were asked the extent to which they had observed the influence of other people's presence in the responses (no influence at all; less than five responses; between six and 10 responses; more than 10 responses). Results show that in nine out of ten cases, enumerators reported no influence at all (3,605 observations). Moreover, within the 429 observations in which there was influence over the interview, less than half involved the "worry" question (45 percent) or the likelihood question (44 percent). Further analyses showed that other people's influence during the interview increases with respondent's age, especially when respondents live in a property owned by a family member.

In conclusion, these findings suggest that appropriate enumerator training can be successful in reducing the influence of other people's presence during the interview. This is achieved by (i) thoroughly explaining the study objectives; (ii) stressing the importance of individual perceptions, which are neither correct nor incorrect; and (iii) kindly asking selected respondents for a quiet and private place to conduct the interview. Further research is needed to assess the size of bias in responses associated with presence of other people during the interview and if such bias is different by gender, tenure, or role in a household.

# 8.0 CONCLUSIONS AND RECOMMENDATIONS

## 8.1 CONCLUSIONS

This study of alternative scales for measuring PTS for land and housing property in four municipalities in Colombia yields the following conclusions:

- **Data quality**: The study finds that both scales produce a similar level of data quality. The predicted value of TPTS explains 89 percent of the variation in the 4-point scale and 93 percent in the 5-point scale. The higher value on the 5-point scale suggests that this scale is able to capture additional variation in the underlying TPTS compared with the 4-point scale.
- **Number of categories:** Looking at the optimal number of categories to best fit the predicted values of TPTS, the study finds that there is no one right or best way to classify the observations. Both four and five categories perform well and determining which produces the best fit depends on the characteristics of the specific sub-population studied. Different sub-populations (e.g., municipalities, holders of different forms of tenure) can be optimally classified into four, five, or six categories. The optimal number of categories for a particular data set depends on proportions of the relevant sub-populations in the data, which is not known to a researcher prior to data collection. For consistency and comparability across countries and institutional contexts, either 4- or 5-point scales can be selected as a standard and consistently applied. However, in cases when comparability is less important, the practicality of scale alternatives in a particular context should be considered on a case-by-case basis.
- **Labeling and data bias:** Looking at which scale produces the best matches in classification between the predicted category of TPTS values and the respective category given on the two scales, the study finds that the 4-point scale classifies 89 percent of observations correctly. On the other hand, the 5-point scale only classifies 33.9 percent of observations correctly, with more incorrect classifications in the middle categories. Greater language complexity on the middle points of the 5-point scale may contribute to the mismatch between true feelings of tenure security and the proposed category labels on the survey question. Redesign of this scale would require (i) simplification of the language used for labeling, and (ii) a change in proportion of positive and negative values to make it closer to the predicted distribution of TPTS (a larger number of positive categories). Translation of such a scale into other languages may change the level of complexity and the proportion of positive and negative values, which may lead to errors and inconsistent results across the cultures and languages and, thus, would require a careful quality assessment. For example, a comparison of the asymmetrical 5-point scale with an alternative symmetrical 5-point scale (used for data collection for the LFP impact evaluation baseline conducted by NORC) covering the same municipalities demonstrates a very different distribution of responses across the 5-point response categories (see Annex 3). In contrast, the 4-point scale is better able to overcome the above difficulties as a shorter scale can rely on simpler language to describe the categories, which is reflected in the lower percentage of misclassified observations.
- **Matching the target population:** The 5-point scale uses language of higher complexity to express nuances in meaning among a larger number of categories, which does not match the levels of language typically used by the survey respondents. For example, some feedback from survey enumerators was that that the labels were "too academic," which is confirmed by the

linguistic analysis described above. A shorter scale requires less complex language and provides a better match across languages and linguistic skill levels.

- **Middle point interpretation:** The presence of a middle point in the 5-point scale creates a risk of misclassification of observations that report the middle category into secure or insecure. This risk is higher in the case that the re-classification assumptions do not match the "true" distribution of perceived tenure security. The 4-point symmetrical scale does not have this risk as the respondent must make an explicit choice between the secure and insecure categories without any assumptions by the data user.
- **Comparability:** The study compares the reclassification of the 4-point, 5-point, and predicted values of TPTS into binary variables of tenure security to assess if this helps reduce bias in the data. Both scales perform well. Compared with the binary classification of TPTS, the 4-point scale produces a 94 percent match in binary classification, and the 5-point scale produces a 99 percent match. Thus, reclassifying data on alternative scales into a binary measure of PTS does seem to allow for a reduction in the bias in estimates of tenure security. However, such gain in comparability comes at a cost of losing information on true variation in the levels of tenure security, which may be particularly useful in assessment of changes in the institutional environment or practices (e.g., when assessing the impact of program interventions or information campaigns). Thus, the binary presentation of PTS should be used to gain comparability across the alternative survey instruments and not for data collection itself.
- **Reliability and item non-response:** Both scales demonstrate similar levels of item non-response and capture by far the largest share of variation in the TPTS among all the considered alternative measures of PTS.
- **Implications:** Our results echo conclusions by Saris et al. (2011) and Oberski et al. (2010) that a cross-country comparison of data on PTS requires minimization of sources of measurement errors and adjustments for a potential difference in data quality. Such adjustments require further research of variation in data quality across countries. The study has compared 13 alternative ways to ask question about perceptions of tenure security. The questions based on the likelihood of losing property rights have demonstrated the highest data quality in terms of correlation with the predicted value of PTS.

In general, this study demonstrates that within a given context of PTS, the difference in data quality across alternative measures of PTS is driven not as much by the number of points on the scale, but by the quality of labels and presence of the middle point. Also, comparability with results from other studies requires adjustments to the same target population (e.g., adult population vs. landowners).

## 8.2 RECOMMENDATIONS

While this experiment is based on data from municipalities affected by recent violent conflicts and the version of 4- and 5-point scale is translated only in one language, Spanish, the results suggest some broader practical recommendations for consistent measurement of PTS at the global and local levels, and for considering PTS in policy evaluations and program design. The results also suggest a need for ongoing research. Our recommendations based on this study are:

- A shorter (4-point) symmetrical scale demonstrates a higher robustness to misclassification and errors in translation and is easier for implementation and conversion into a binary scale. Therefore, it is recommended for assessment of PTS across countries and institutional environments. The 5-point scale may be more practical in cases where the measurement needs to be more sensitive to small changes in PTS (e.g., in impact evaluation). However, the design of

labels for a 5-point scale needs more consideration. A scale with a larger number of positive (secure) categories than negative is likely to demonstrate better data quality results.

- Additional studies are necessary to check the robustness of this recommendation to other languages and institutional contexts. Consistency of the global assessment of tenure security will be facilitated by development and testing of standard labels for tenure security scales in multiple languages with a clear guidance for categorization into the binary comparison of secure or insecure tenure.
- A binary measure of tenure security should be used for comparison across countries regardless of the design of original scale used for the data collection (OECD, 2013). The quality of such a binary measure will be conditional on the clarity and consistency of instructions for constructing the binary measure and the original question used for assessment of PTS.

## 8.3 FURTHER RESEARCH

Development of comparable and consistent measures of PTS and instructions for their construction and use in policy analysis and program design would benefit from further research. Some directions for such research include the following:

- Assessment of PTS data quality across languages, countries, institutions, and tenure types (e.g., using Prindex data from 140 countries);
- Assessment of the quality of alternative labelling approaches (e.g., labelling only the end points, using additional visualization materials);
- Follow up study on differences in PTS between the treatment and control LFP communities at the baseline and endline of the intervention; and
- Development of alternative non-survey instruments for assessment of tenure security and assessment of the quality of such instruments.

# REFERENCES


Althauser, R. P., Heberlein, T. A., & Scott, R. A. (1971). A causal assessment of validity: The augmented multitrait-multimethod matrix. In H. M. Blalock, Jr. (Ed.), *Causal models in the social sciences* (pp. 151-169). Chicago: Aldine.

Alwin, D. F. (2007). *Margins of error: A study of reliability in survey measurement.* Hoboken, New Jersey: Wiley and Sons, Inc.

Alwin, D. F. (2010). How good is survey measurement? Assessing the reliability and validity of survey measures. In P. V. Marden & J. D. Wright (Eds.), *Handbook of survey research.* Emerald Publishing Limited.

Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement. *Sociological Methods and Research, 20*, 139-181.

Barthel, K., Cespedes, V., Salazar, B., Torres, R., & Varón, M. (2016). Land and rural development policy reforms in Colombia: The path to peace. *Annual World Bank Conference on Land and Poverty 2016.* https://www.land-links.org/wp content/uploads/2016/09/USAID_Land_Tenure_World_Bank_2016_LRDP_Policy_Reforms.pdf

Barthel, K., Childress, M., Magen, E., Nizalov, D., Gaviria, A., & Varela, D. (2021). *Prindex baseline data and final analysis report.* Bogota: USAID Land for Prosperity Activity.

Besley, T. (1995). Property rights and investment incentives: Theory and evidence from Ghana. *Journal of Political Economy, 103*(5), 903-937

Coenders, G., & Saris, W. E. (1995). 'Categorization and quality: The choice between Pearson and Polychoric correlations." In W. E. Saris & A. Munnich (Eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments* (pp. 125-144). Budapest: Eotvos University Press.

DANE (Departamento Administrativo Nacional de Estadística). (2018). *Censo Nacional de Vivienda y Población*. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/informacion-tecnicahttps://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/informacion-tecnica

Deininger, K. (2003). Land policies for growth and poverty reduction, *World Bank Policy Research Report series*. Washington, D.C.: World Bank; Oxford and New York: Oxford University Press

FAO, World Bank, & UN-Habitat. (2019). *Measuring individuals' rights to land: An integrated approach to data collection for SDG Indicators 1.4.2 and 5.a.1*. Washington, DC: World Bank. http://www.fao.org/documents/card/en/c/ca4885en/

Feyertag, J., Childress, M., Flynn, R., Langdown, I., Locke, A. & Nizalov, D. (2020). *Prindex comparative report: A global assessment of perceived tenure security from 140 countries*, https://www.prindex.net/reports/prindex-comparative-report-july-2020/

Fowler Jr, F. J., & Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Sage.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 34,* 409-426.

Juboori, S. I., & Muhaisin, M. H. (2017). Niveles de uso del lenguaje y sus tipos. *Al-Adab Journal*, *121*, 1438. https://www.iasj.net/iasj/download/e09bf8f21cc7c5bc

Kantardzic, M. (2019). *Data mining: Concepts, models, methods, and algorithms*. Wiley.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Mardsen & J. D. Wright (Eds.), *Handbook of survey research*. Emerald Publishing Limited.

Lawry, S., Samii, C., Hall, R., Leopold, A., Hornby, D. & Mtero, F. (2017). The impact of land property rights interventions on investment and agricultural productivity in developing countries: a systematic review. *Journal of Development Effectiveness, 9*(1), 61-81

Liao, P., Saris, W. E., & Zavala-Rojas, D. (2019). Cross-national comparison of equivalence and measurement quality of measurement scales in Denmark and Taiwan. *Journal of Official Statistics*, *35*(1), 117-135

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *WIREs: Data Mining and Knowledge Discovery, 2*(1), 86-97.

Nizalov, D., Langdown, I., Feyertag, J., Childress, M., & Locke, A. (forthcoming). De-facto vs. de-jure security of property rights: evidence from the cross-country Prindex survey. Under review by *Journal of Land Economics*.

Oberski, D., Saris, W. E., & Hagenaars, J. (2010). Categorization errors and differences in the quality of questions in comparative surveys. In *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 435-453). Wiley.

OECD. (2013). *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing. http://dx.doi.org/10.1787/9789264191655-en

Persha, L., Haugan, G., Mittelberg, T., Wendt, R., & Protik, A. (2020). *Evaluation of the 'Supporting Deforestation-Free Cocoa in Ghana' Project Bridge Phase: Baseline Report.* Washington, DC: USAID Communications, Evidence and Learning (CEL) Project. https://www.land-links.org/wp-content/uploads/2020/06/CEL-Ghana_REVISED-Baseline-Report_NORC_5may2020_CLEANv2.pdf

Prindex. (2018). *PRIndex Baseline Survey_Final.* https://www.prindex.net/documents/66/Prindex_2018_questionaire.pdf

Revilla, M., & Saris, W. E. (2013). The split-ballot multitrait-multimethod approach: Implementation and problems. *Structural Equation Modeling: A Multidisciplinary Journal, 20,* 27-46. https://doi.org/10.1080/10705511.2013.742379.

Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree-disagree scales. *Sociological Methods & Research, 43*(1), 73-97. https://doi.org/10.1177/0049124113509605

Saris, W. E., & Gallhofer. I. N. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods, 1,* 29-43. http://dx.doi.org/10.18148/srm/2007.v1i1.49

Saris, W. E., & Gallhofer, I. N. (2014). *Design, evaluation, and analysis of questionnaires for survey research*. John Wiley & Sons.

Saris, W., Oberski, D., Revilla, M., Zavala, D., Lilleoja, L., Gallhofer, I., & Gruner, T. (2011). The development of the program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working Paper* 24, Universitat Pompeu Fabra. https://www.upf.edu/documents/3966940/3986764/RECSM_wp024.pdf.

Saris, W. E., Satorra, A., & Coenders, G. (2004). A new approach to evaluating the quality of measurement instruments: The split-ballot MTMM design. *Sociological Methodology*, *34*(1), 311-347. https://doi.org/10.1111/j.0081-1750.2004.00155.x

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236-244.

Werts, C. E., & Linn, R.L. (1970). Path analysis. Psychological examples. *Psychological Bulletin, 74*, 193–212.

# ANNEX 1. ESTIMATION RESULTS AND STATISTICAL ANALYSIS

## TABLE A1.1. ESTIMATION RESULTS FOR THE STRUCTURAL EQUATION MODEL

| Model: | | Latent variables and path coefficients | | |
|---|---|---|---|---|
| # | Dependent variable | TPTS | Pt4 | W4pt |
| 1 | Worry question | 1 | 0.0462 | 0.349*** |
| | | (0) | (0.238) | (0.044) |
| 2 | Likely question – 4 Pt | 2.143 | 1 | |
| | | (1.881) | (0) | |
| 3 | Likely question - 5 Pt | 2.004 | | |
| | | (1.279) | | |
| 4 | Competing claim | 0.929*** | -1.525 | 1 |
| | | (0.147) | (1.061) | (0) |
| 5 | Local authority dispute | 0.632*** | -1.026 | 0.982*** |
| | | (0.111) | (0.783) | (0.080) |
| 6 | Family job loss | 0.575*** | 0.307 | 0.798*** |
| | | (0.078) | (0.517) | (0.108) |
| 7 | Family dispute | 0.569*** | 0.155 | 0.795*** |
| | | (0.083) | (0.542) | (0.072) |
| 8 | Fraud | 0.752*** | -1.689 | 0.959*** |
| | | (0.118) | (1.125) | (0.061) |
| 9 | Company taking | 0.727*** | -2.248 | 1.033*** |
| | | (0.123) | (1.372) | (0.082) |
| 10 | Boundary dispute | 0.590*** | -0.774 | 0.777*** |
| | | (0.095) | (0.584) | (0.059) |
| 11 | General protection | -0.144*** | 0.28 | |
| | | (0.050) | (0.219) | |
| | | | | |
| | Observations | 4,033 | var(PTS) | 1.586** |
| | LL | -1565000 | | (0.705) |
| | | | var(Pt4) | 0.534 |
| | | | | (0.647) |
| | | | var(W4pt) | 7.294*** |
| | | | | (1.361) |

*Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1*

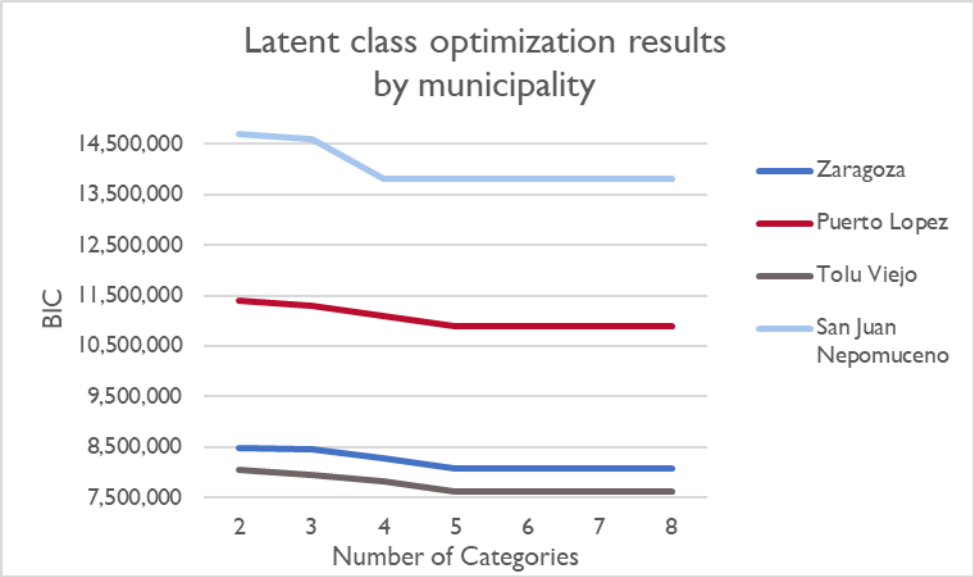
## TABLE A1.2. EFFECT OF THE COMMON INSTRUMENT DESIGN FEATURES ON PREDICTED TPTS

| Variables | Predicted TPTS |
|---|---|
| Worry (predicted instrument effect) | 0.078*** |
| | (0.010) |
| 4-point (predicted instrument effect) | 0.310*** |
| | (0.053) |
| Constant | 0.032 |
| | (0.023) |
| Observations | 4,034 |
| R-squared | 0.036 |

*Note: estimated with OLS; sampling weights are applied; Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1*

# ANNEX 2. ROBUSTNESS OF TPTS CLASSIFICATION RESULTS

### FIGURE A2.1. LATENT CLASS ANALYSIS BY MUNICIPALITY

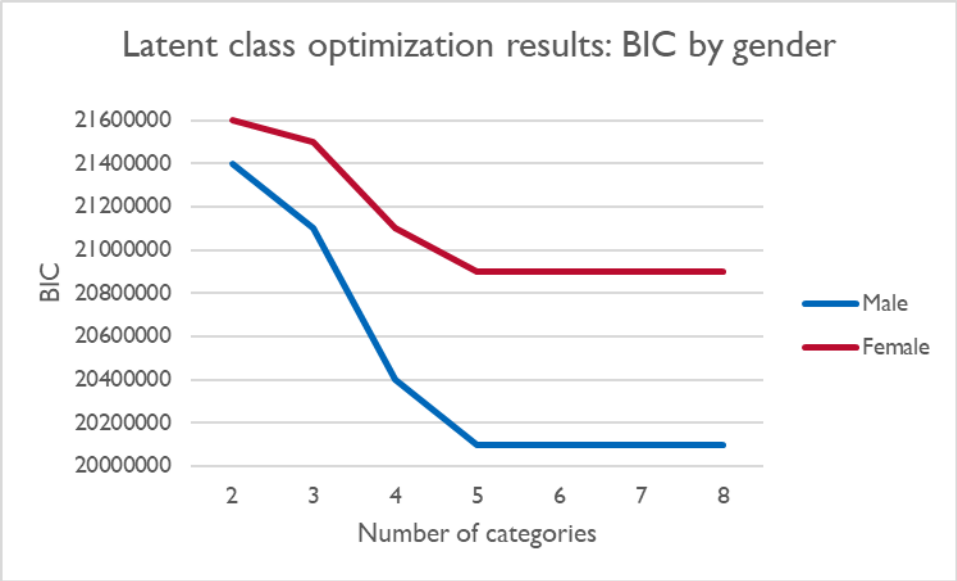



*Note: sampling weights are applied*

### TABLE A2.1. TENURE SECURITY CLASSIFICATION ACROSS MUNICIPALITIES

| Municipality | Freq. | TPTS mean value | TPTS (secure) | Worry (secure) | Likely (secure) | Categories |
|---|---|---|---|---|---|---|
| Zaragoza | 1,015 | 0.13 | 60.2% | 31.7% | 61.9% | 6 |
| Puerto Lopez | 1,013 | 0.01 | 61.1% | 38.4% | 60.0% | 5 |
| Tolu Viejo | 1,011 | 0.02 | 68.5% | 30.4% | 71.5% | 5 |
| San Juan Nepomuceno | 995 | -0.02 | 66.2% | 42.4% | 66.2% | 4 |
| **Total** | **4,034** | **0.03** | **64.2%** | **36.9%** | **64.8%** | |

## FIGURE A2.2. LATENT CLASS ANALYSIS BY GENDER

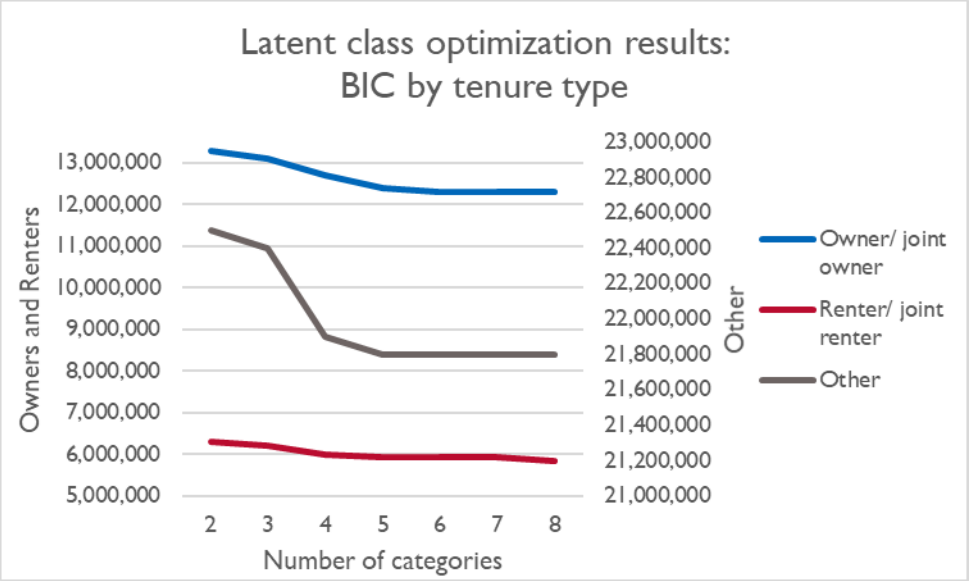



*Note: sampling weights are applied*

## TABLE A2.2. TENURE SECURITY CLASSIFICATION BY GENDER

| Gender | Freq. | TPTS mean value | TPTS binary | Worry (secure) | Likely (secure) | Categories |
|---|---|---|---|---|---|---|
| Male | 1,957 | 0.01 | 64.9% | 37.8% | 66.0% | 5 |
| Female | 2,077 | 0.05 | 63.5% | 35.9% | 63.5% | 5 |
| **Total** | **4,034** | **0.03** | **64.2%** | **36.9%** | **64.8%** | |

## FIGURE A2.3. LATENT CLASS ANALYSIS BY TENURE TYPE

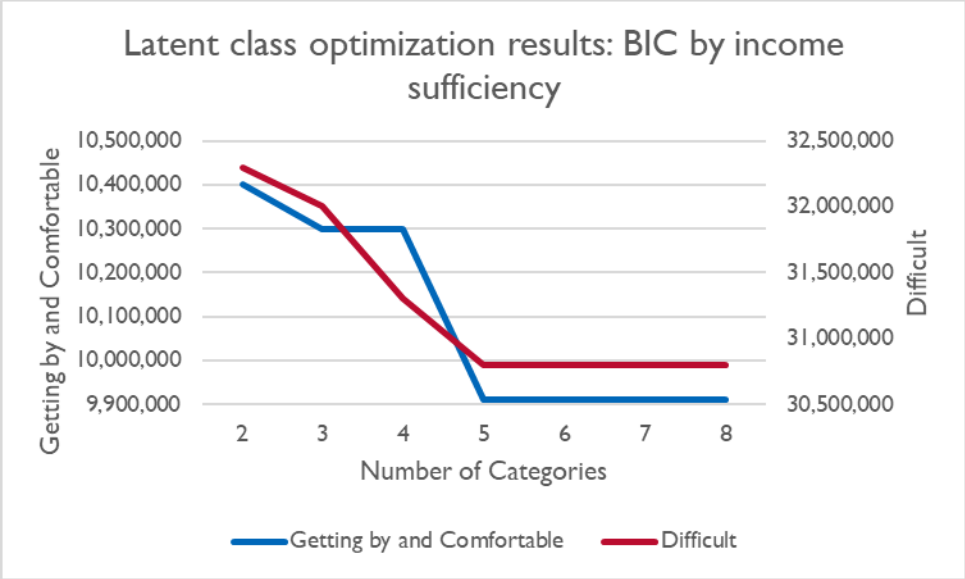



*Note: sampling weights are applied*

## TABLE A2.3. TENURE SECURITY CLASSIFICATION BY TENURE TYPE

| Tenure | Freq. | TPTS | TPTS binary | Worry (secure) | Likely (secure) | Categories |
|---|---|---|---|---|---|---|
| Owner/joint owner | 1,451 | -0.241 | 76.8% | 43.2% | 79.2% | 6 |
| Renter/joint renter | 531 | 0.349 | 45.9% | 31.4% | 44.4% | 8 |
| Other | 2,052 | 0.108 | 61.1% | 34.4% | 61.2% | 5 |
| **Total** | **4,034** | **0.026** | **64.2%** | **36.9%** | **64.8%** | |

## FIGURE A2.4. LATENT CLASS ANALYSIS BY INCOME SUFFICIENCY




*Note: sampling weights are applied*

## TABLE A2.4. TENURE SECURITY CLASSIFICATION BY INCOME SUFFICIENCY

| Income sufficiency | Freq. | TPTS mean value | TPTS binary | Worry (secure) | Likely (secure) | Categories |
|---|---|---|---|---|---|---|
| Getting by and comfortable | 1,029 | -0.191 | 73.4% | 46.0% | 73.5% | 5 |
| Difficult | 3,005 | 0.101 | 61.0% | 33.7% | 61.8% | 5 |
| **Total** | **4,034** | **0.026** | **64.2%** | **36.9%** | **64.8%** | |

# ANNEX 3. PTS OF OWNERS IN ILRG SAMPLE WITH COMPARISON TO LFP IMPACT EVALUATION BASELINE FINDINGS

Table A3.1 and Figure A3.1 demonstrate the distribution of PTS responses by land property owners measured with three alternative scales: 4-point and asymmetrical 5-point scales used in this ILRG-funded Prindex study and the 5-point symmetrical scale used in the baseline survey for the LFP impact evaluation (IE) being implemented by NORC. The data compared in this annex were collected in the same municipalities during the same time period and the samples are representative of the municipal population. We can clearly observe the difference in distribution of observations across the individual categories of both versions of 5-point scales, which indicates the impact of alternative sets of labels.

### TABLE A3.1. PTS DISTRIBUTION FOR OWNERS

| ILRG: 4-point | | ILRG: 5-point | | LFP IE: 5-point | |
|---|---|---|---|---|---|
| **Scale Labels** | **Percent** | **Scale Labels** | **Percent** | **Scale Labels** | **Percent** |
| Very unlikely | 46.7% | Not at all likely | 39.4% | Not at all likely | 70.30% |
| Unlikely | 34.8% | Slightly likely | 37.4% | Somewhat unlikely | 7.30% |
| Somewhat likely | 10.9% | Moderately likely | 14.9% | Neutral | 3.10% |
| Very likely | 5.1% | Very likely | 6.3% | Somewhat likely | 15.90% |
| | | Extremely likely | 0.7% | Extremely likely | 3.10% |
| *Don't know* | *2.1%* | *Don't know* | *1.1%* | *Don't know* | *0.30%* |
| *Refused* | *0.4%* | *Refused* | *0.2%* | *Refused* | *0.02%* |

### FIGURE A3.1. HISTOGRAMS FOR 5-POINT ILRG AND LFP IE SCALES

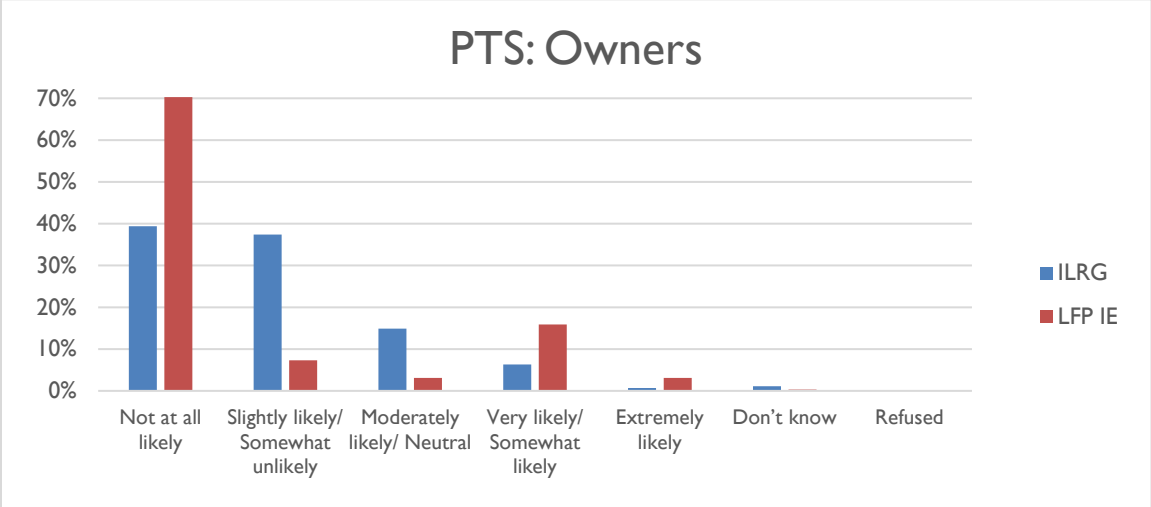



Table A3.2 demonstrates the values of binary measures of PTS for property owners constructed based on the above-mentioned alternative data sets. Comparison of the results for two versions of 5-point scales when converted into the binary scale demonstrate similarity in estimates of the preferred indicators for security (S1) and insecurity (I1) of tenure. For construction of the preferred measures out of symmetrical 5-point scale, the middle value was allocated to the insecure category.

## TABLE A3.2. ALTERNATIVE MEASURES OF PTS FOR OWNERS

| Measure | ILRG: 4-point scale | | ILRG: 5-point scale | | LFP IE: 5-point scale | |
|---|---|---|---|---|---|---|
| | Construction | Statistics, % | Construction | Statistics, % | Construction | Statistics, % |
| Secure tenure (s1) | [(1) Very unlikely + (2) Unlikely)] / Total | 81.5% (2.5%) | [(1) Not at all likely+ (2) Slightly likely] / Total | 76.8% (2.1%) | [(1) Not at all likely+ (2) Somewhat unlikely] / Total | 77.6% |
| Secure tenure (s2) | (1) Very unlikely /Total | 46.7% (2.8%) | (1) Not at all likely / Total | 39.4% (2.5%) | (1) Not at all likely / Total | 70.3% |
| Insecure tenure (i1) | [(3) Somewhat likely + (4) Very likely] / Total | 16.0% (2.4%) | [(3) Moderately likely + (4) Very likely + (5) Extremely likely] / Total | 21.9% (2.1%) | [(3) Neutral + (4) Somewhat likely + (5) Extremely likely] / Total | 22.10% |
| Insecure (extremely) tenure (i2) | (4) Very likely / Total | 5.1% (1.5%) | [(4) Very likely + (5) Extremely likely] / Total | 7.0% (1.5%) | [(4) Somewhat likely + (5) Extremely likely] / Total | 19.0% |
| Insecure (extremely) tenure (i3) | (4) Very likely / Total | 5.1% (1.5%) | (5) Extremely likely / Total | 0.7% (0.3%) | (5) Extremely likely / Total | 3.1% |

*Note: std. err. in parenthesis*

Table A3.3 shows the values of the preferred binary measures of PTS constructed out of the ILRG data separately for owners, renters, and users of family property. Only in one case is a statistically significant difference (at 10 percent significance level) detected among the estimates produced out of 4-point and 5-point original measures of PTS, demonstrating that conversion to the binary scale helps to establish comparability among the two considered alternatives. However, a significant difference in the levels of tenure security and insecurity is demonstrated across the forms of tenure. This result implies that focusing analysis on tenure security of property owners yields results which cannot be extended to other parts of the population. It also implies that comparability across alternative measures of PTS should be established for comparable populations (e.g., in terms of forms of tenure).

## TABLE A3.3. PREFERED BINARY MEASURES OF PTS ACROSS THE FORMS OF TENURE

| Sample | Secure tenure (s1) | Insecure tenure (i1) |
|---|---|---|
| **Owners** | | |
| 4-point scale | 81.5% | 16.0% |
| 5-point scale | 76.8% | 21.9% |
| Difference/ Significance | 4.8% | -5.89%* |
| **Renters** | | |
| 4-point scale | 43.0% | 55.4% |
| 5-point scale | 49.3% | 48.4% |
| Difference/ Significance | -6.3% | 7.0% |
| **Family users** | | |
| 4-point scale | 63.7% | 34.6% |
| 5-point scale | 63.4% | 35.8% |
| Difference/ Significance | 0.3% | -1.2% |

*Note: *** p<0.01, ** p<0.05, * p<0.1 for differences*

**U.S. Agency for International Development**
1300 Pennsylvania Avenue, NW
Washington, DC 20523
Tel: (202) 712-0000
Fax: (202) 216-3524
**www.usaid.gov**