



A TOURIST DEVELOPMENT IN MA'AN GOVERNORATE BY JACOB PATTERSON-STEIN

CITIES EVALUATION BASELINE REPORT

MARCH 2019

This publication was produced for review by the United States Agency for International Development. It was prepared by Dan Killian, Dima Toukan, Jacob Patterson-Stein and Nikki Zimmerman for Management Systems International (MSI), A Tetra Tech Company.

CITIES EVALUATION BASELINE REPORT

Cities Implementing Transparent, Innovative, and Effective Solutions (CITIES)

Contracted under AID-278-C-13-00009

USAID/Jordan Monitoring and Evaluation Support Project (MESP)

DISCLAIMER

The authors' views expressed in this report do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

CONTENTS

ACRONYMS	II
EXECUTIVE SUMMARY	I
INTRODUCTION	6
BACKGROUND	7
CITIES ACTIVITY OVERVIEW	9
THEORY OF CHANGE	10
EVALUATION QUESTIONS	12
EVALUATION DESIGN	14
EVALUATION INDICATORS	14
MATCHING FOR COMPARISON GROUP CONSTRUCTION	16
QUANTITATIVE DATA COLLECTION & ANALYSIS APPROACH	17
GENERAL POPULATION SURVEY	19
IMPLEMENTATION FIDELITY	19
LEARNING AGENDA	20
EVALUATION LIMITATIONS	20
BASELINE DATA COLLECTION	23
QUANTITATIVE DATA COLLECTION	23
QUALITATIVE DATA COLLECTION	25
KEY FINDINGS	26
EVALUATION QUESTION 1: EFFECTIVENESS	26
EVALUATION QUESTION 2: SUSTAINABILITY	36
EVALUATION QUESTION 3: SYNERGY	38
INTERNAL VALIDITY OF MICA SCORES	41
ASSESSING BASELINE EQUIVALENCE	45
IMPLEMENTATION FIDELITY	60
PUBLIC PERCEPTION INDICATORS - GENERAL POPULATION SURVEY	62
CONCLUSIONS	69
RECOMMENDATIONS	72
ANNEXES	74
ANNEX 1: NOTES ON IMPLEMENTATION FIDELITY	74
ANNEX 2: INVESTIGATING SCORER AGREEMENT	84
ANNEX 3: INTER RATER AGREEMENT, BY MUNICIPALITY	95
ANNEX 4: INVESTIGATING BASELINE EQUIVALENCE	109
ANNEX 5: GENERAL POPULATION SURVEY	115
ANNEX 6: ANALYZING INSTITUTIONAL CAPACITY	120
ANNEX 7: LISTING OF SELECTED TREATMENT AND COMPARISON MUNICIPALITIES	123
ANNEX 8: QUALITATIVE DATA COLLECTION GUIDES	125
ANNEX 9: RELEVANT LEARNING AGENDA FINDINGS	131

ACRONYMS

AECID	Spanish Agency for International Development Cooperation
AFD	French Development Agency
CBO	Community Based Organizations
CEP	Community Engagement Project
CITIES	Cities Implementing Transparent, Innovative, and Effective Solutions
d-i-d	Difference in differences
ESSRP	Emergency Services and Social Resilience Project
ET	Evaluation Team
EQ	Evaluation Question
FCM	Federation of Canadian Municipalities
FGD	Focus Group Discussion
FY	Fiscal Year
GESI	Gender and Social Inclusion
GoJ	Government of Jordan
ICC	Intra-cluster correlation coefficient
IE	Impact evaluation
KII	Key Informant Interview
LDU	Local Development Unit
MEL	Monitoring, Evaluation, Learning
MENA	Middle East and North Africa
MESP	Monitoring and Evaluation Support Project
MoI	Ministry of Interior
MAP	Municipal Action Plan
MICA	Municipal Institutional Capacity Index
MoMA	Ministry of Municipal Affairs
MoPIC	Ministry of Planning and International Cooperation

MSI	Management Systems International
MSSRP	Municipal Services and Social Resilience Project
NGO	Non-governmental organization
OECD	Organization of Economic Cooperation and Development
PSM	Propensity Score Matching
SDIP	Strategic Development Improvement Plan
SOW	Statement of Work
SDIP	Service Delivery Improvement Plan
SML	Sector Municipal Loan
SWM	Solid Waste Management
UNDP	United Nations Development Programme
USAID	U.S. Agency for International Development
VNG	International Cooperation Agency of the Association of Netherlands Municipalities
WB	World Bank

EXECUTIVE SUMMARY

USAID/Jordan has requested technical support from its institutional contractor, the Monitoring and Evaluation Support Project (MESP), in designing a prospective evaluation and learning agenda of its municipal governance initiative, Cities Implementing Transparent, Innovative, and Effective Solutions (CITIES). This evaluation will contribute to an emerging body of evidence in order to help USAID better tailor subsequent democracy and governance investments in Jordan and elsewhere. The findings from this evaluation and the learning agenda will also help USAID, the Government of Jordan, and the broader donor and governance community better understand the impact of decentralization efforts at the municipal institution level and for the citizens who ultimately benefit from efforts to make government more effective, equitable, and responsive.

CITIES is a five-year activity implemented by a consortium led by Chemonics International with an estimated budget of \$58,549,993 and period of performance from September 25, 2016 – September 24, 2021. CITIES aims to increase the effectiveness of municipal governance and support decentralization in Jordan and is implemented in 33 municipalities across all 12 governorates in Jordan. By strengthening municipal institutional capacity, improving public service delivery, enhancing municipal-community communications and strengthening community resilience and cohesion, it is expected that citizens will ultimately view their local municipality to be more responsive to their needs and more effective in executing their core functions.

EVALUATION QUESTIONS

The evaluation team (ET) set out to investigate the overall impact of CITIES, and, to complement this line of inquiry, organized its work according to broad performance evaluation questions that examine activity effectiveness, sustainability, synergies, and learning. Evaluation questions include:

1. **Effectiveness:** What is the overall effectiveness of CITIES for achieving its goals of enhancing democratic governance, improving trust and confidence in elected officials, and strengthening stability?
 - a. Did the activity achieve its intended outcomes?
 - b. Are there certain areas/activities and components that have been more effective than other areas/activities and components?
2. **Sustainability:** Which interventions under CITIES are most likely to be sustained over time? Why and how? To what extent are CITIES capacity building efforts being sustained by municipalities and other targeted bodies of decentralization?
3. **Synergy:** How well has CITIES developed synergies (coordination and collaboration) with other USAID activities that focus on similar objectives, for example, decentralization and working with Municipalities? Are donors operating in the decentralization and local governance space achieving synergies through coordination or knowledge sharing?
4. **Learning:** Based on the performance of CITIES, both in terms of effectiveness and sustainability, what are some key lessons learned, by component and sub-activity?

EVALUATION DESIGN AND LIMITATIONS

The evaluation utilizes a mixed-methods, quasi-experimental design approach that focuses on both impact and performance aspects. Key aspects of the design include quantitative institutional measures between 15 treatment and 15 comparison municipalities (as measured by the CITIES Municipal Institutional Capacity Assessment (MICA) tool) to measure changes in municipal capacity; qualitative interviews with key informants on state and local governance and decentralization, as well as small-scale surveys of direct beneficiaries of municipal service improvements (planned for mid-line data collection) to inform contextual understanding and performance related evaluation questions; and a large-scale General Population household survey to assess public perception of municipal responsiveness and effectiveness. In addition, the evaluation also focuses on tracking implementation fidelity, as well as a series of learning agenda questions to inform overall findings, conclusions, recommendations, and broader learning goals.

Key limitations of this design include:

Unobservable factors and data scope: The MICA tool was designed to help CITIES inform and target its implementation but may not fully capture municipal performance (e.g. factors that affect MICA scores that are not part of CITIES's implementation, or factors not captured by MICA).

External validity: Jordan has undergone multiple institutional reform efforts and it is likely that changes to the local governance context will continue to evolve. Due to these contextual factors, results from this evaluation may not hold in a future context or even at a different governing-level.

Changing implementation context: CITIES implementation approach is responsive to local needs, which is valuable for beneficiaries as it likely increases the relevance of activities. If interventions are provided that, for example, do not directly affect MICA scores it may be the case that no impact is found when positive changes did in fact occur.

Other donor activity: The evaluation will be able to control for the biggest actors, such as the Emergency Services and Social Resilience Project (ESSRP)/ Municipal Services and Social Resilience Project (MSSRP) but will not be able to track all donor activity that may also affect CITIES outcomes. If untracked donor activity realizes substantive outcomes, it will distort the CITIES impact estimates in ways that depend on the distribution of that donor activity.

BASELINE DATA COLLECTION

To measure across evaluation indicators, MESP collected data utilizing both quantitative and qualitative data collection as part of the baseline. Overall, through baseline data collection efforts, MICA data was collected from 48 municipalities (33 treatment municipalities collected through CITIES, 15 comparison municipalities collected by the ET), 20 key informant interviews (KIIs) with key stakeholders in comparison and treatment municipalities, and completion nearly 12,000 interviews at the household level as part of the General Population Survey.

This baseline report, a) summarizes qualitative inquiry in a selection of municipalities on the current state of municipal governance and decentralization, b) assesses the internal validity of MICA measures in comparison municipalities, such that they may be considered comparable to CITIES' MICA measurements in treatment municipalities, and c) in comparing treatment and comparison MICA measures, assesses

whether there is sufficient pre-treatment equivalence across treatment and comparison municipalities to enable robust inference on CITIES impact at the institutional municipal level.

KEY FINDINGS AND CONCLUSIONS

These baseline findings highlight the failings of any one approach and the imperative to look across methods. Valid and defensible impact estimates based on institutional capacity scores must rely on a combination of inferential analysis of MICA scores and thick description of what actually changes in municipal capacity. Moving from municipal institutional capacity to citizen perception of municipal effectiveness, achieving citizen engagement and service provision improvement involve complex political, financial, and social systems. The qualitative and quantitative findings presented here provide a strong overall baseline for assessing changes over time and generating insight and learning for future program design.

EVALUATION QUESTIONS

EQ1: Effectiveness

Municipalities suffer from significant structural burdens that impair the provision of quality services to citizens. The effectiveness of the CITIES project will depend on its ability to navigate the change that can reasonably be expected within this new decentralized system. The most pronounced challenges are those related to municipalities' limited financial resources and revenue streams, the encroachment of the central government on their autonomy and independence and their overall weak institutional capacity. Important findings include:

- Municipal capacity to strengthen local service delivery is limited overall due to limited autonomy, staffing, and resource constraints in both assignment groups;
- Community engagement mostly relies on direct contact and personal relations. There is lack of systematic feedback or use of citizens input for planning, and there are fears of raising expectations that municipalities cannot meet;
- Local councils have no resources and an unclear role across both assignment groups. There is almost no reporting happening between mayors and the municipal councils in the comparison group;
- The governorate council is still determining its role and relationship with the central ministries and municipalities, with respondents noting the limited devolution of power, the need for legislative reforms and policy guidance to implement the decentralization law;
- There remains confusion around decentralization, gaps in the legal framework, weak engagement and linkages between governmental layers, and continued centralized decision-making. Respondents report mixed perceptions of the effectiveness of CITIES's support for capacity building around decentralization.

EQ2: Sustainability

More time will be needed to determine the sustainability of specific CITIES interventions, but a number of factors and risks to sustainability emerged. These include:

- Endorsement of CITIES' system level initiatives by the national government
- Institutionalization of decentralization

- Municipal capacity constraints and control
- Harmonized donor capacity building support strategies
- Civil society capacity to fulfil its role in municipal accountability
- Public perception and heightened expectations of municipalities beyond what municipalities can deliver

EQ3: Synergy

Baseline data suggests there has been some initial coordination between CITIES and other donor activities, but there are salient gaps in overall donor coordination and synergy to overcome. Through this baseline study, over 28 implementing agencies operating through 48 projects implementing 208 separate interventions at the central and local levels was identified. Beyond CITIES-led coordination, donors agree on the existence of an overall national vision for change but emphasized that a clear roadmap for how to bring this change about is still lacking.

INTERNAL VALIDITY AND BASELINE EQUIVALENCE

Comparison MICA scores are generally considered internally valid among MESP raters and with CITIES scorers, but some degree of inherent subjectivity remains. Measurements of institutional capacity have intrinsic uncertainty, but the ET has been vigilant in trying to minimize the error that comes from differences in interpretation. There are some measures, and some municipalities, where error is relatively high. This error presents risks to the impact estimates comparing individual municipalities, but in aggregate the error is tolerable and does not limit the evaluation design.

While the reduction in scoring error reduced this overall measurement error, it presents the need to be cautious in inferential analysis. To address this concern, the evaluation team concluded that sensitivity analysis of equivalency using multiple configurations across matched, partially matched, and the full CITIES sample are needed to assess the validity of estimates.

There is baseline equivalency across treatment and comparison municipalities at the aggregate level for MICA service scores, but there is a much greater degree of variability at the individual item level. MICA scores are organized by service delivery (services) and gender and social inclusion (GESI) metrics. After reviewing two sets of scores across six configurations, the evaluation team found that the distribution of services scores at the aggregate level are balanced across assignment groups for matched (15-15) and partially matched (15-23) municipalities for MICA service scores, but not sufficiently balanced for GESI scores. This implies that on average the CITIES and comparison municipalities are statistically similar for MICA services scores. At the MICA element/item level, due to a higher degree of item-level variability, inference at this level will require triangulation with other data streams. Overall, our analysis of equivalency provides hope that any measurement of difference at endline can be attributed to CITIES rather than unobserved characteristics or is a result of selection bias.

IMPLEMENTATION FIDELITY

Implementation of CITIES interventions prior to the evaluation baseline will not affect the design or implementation of the evaluation. The CITIES evaluation has been carefully designed around the schedule of CITIES's inputs and work plan. As occurs with many donor initiatives, there has been some expansion of scope, with stakeholders making requests outside of the original work plan or expressing a desire to spend more.

Based on the timeline of implementation, the ET concludes that any activity implemented prior or during the baseline assessment will not affect the design or implementation of the evaluation as they were either unlikely to have a measurable impact on the outcomes of interest in and of themselves or were provided across assignment groups. That said, it is important to continue to track such activities and to maintain communication with CITIES to determine whether and how these requests affect planned activities and resource allocation.

GENERAL POPULATION PERCEPTIONS

In reviewing the schedule of activities and how faithfully they have been implemented to date, it is clear that donors should not expect a dramatic jump in citizen perception. Citizen perception of government has a complex and non-linear relationship with the actual performance of government institutions. CITIES is positioned to help municipalities realize changes institutionally and capacity-wise, but it remains to be seen how this will be perceived by constituents.

RECOMMENDATIONS

Strengthening Implementation Effectiveness and Sustainability:

- CITIES should continue to focus more on contextually nuanced and technical training than awareness-raising efforts especially when targeting municipal staff and leadership.
- To the extent possible, CITIES should tie grants to needs identified in the needs manual sanctioned by official process.
- CITIES should support the unification of all capacity building efforts under one institutional umbrella whether it is MoMA's training directorate or another institution to increase sustainability and advocate that such efforts are sequenced to ensure effectiveness and knowledge retention and application.

Synergy:

- USAID and CITIES should continue and increasingly take an active role to strengthen inter-donor coordination related to support for municipal services and decentralization.
- CITIES and the next ET should support efforts to capture learning generated by the multiple donors in the same development space as CITIES, and its application.

Contextual Understanding:

- CITIES and the next ET should track the following to inform understanding of project context:
 - Changes in legislative and regulatory framework;
 - Regulatory, relational and technical dimensions of the various municipal functions;
 - Relational dynamics between municipal leadership staff including mayors, executive directors and local and municipal council members;
 - Automation process in municipalities;
 - Resource allocation to municipalities and governorates and related decision-making processes; and
 - The National Strategy of Decentralization and the overall direction of decentralization.

Implementation Fidelity Tracking:

- CITIES should develop a standardized component-level or project-wide tracking system for monitoring activities by date, type, and location. This will help improve implementation, measure scope creep, and provide inputs for further internal and external study.

Evaluation Implementation:

- Considering reliability issues with municipal contextual indicators such as salaries and budget expenditure, the ET should confirm values with national ministries during future data collection rounds.
- Considering that in comparison municipalities that had GESI items present, the median services score was slightly higher than that of comparison municipalities that did not have any GESI items present, the ET should look into conducting additional analysis and qualitative study to understand the relationship between the GESI and services scores.
- The ET and CITIES should maintain or even expand the current level of coordination across the implementation and evaluation teams. The ET used the CITIES' MICA tool as a data collection instrument, which required training and trust, and supported the ET to mitigate rater error. To support this, USAID should continue to support the open communication between the ET and IP stakeholders.
- The ET and CITIES should consider holding additional discussions before the next round of data collection to ensure the ET is properly aligned with the MICA methodology, and to assess the possible use of other CITIES diagnostic tools per CITIES' recommendation.
- The ET should generate impact estimates across both "ET Final MICA Scores" and "Alternate MICA Scores." While the alternate scores are considered the most valid scores, any genuine and durable treatment effect should persist across both sets of estimates.
- Even though comparison MICA scores are generally considered internally valid among MESP raters and with CITIES scorers, the ET should conduct sensitivity analysis of equivalency using multiple configurations across matched, partially matched, and the full CITIES sample to assess the validity of estimates.
- The ET should continue to track donor activity until the next round of MICA data collection is implemented, and incorporate that data into the impact estimates to help isolate the effect of CITIES from the potential effect of other activities.

INTRODUCTION

For much of its history, USAID has relied upon local governance and community development programming as a primary tool to foster democratic norms and institutions. Complementary to this effort has been longstanding support of and advocacy at the central level for political and fiscal decentralization. While community development and political and administrative decentralization has a long history amidst a general consensus that such programming is effective at delivering resources to the local level while building democratic capacity through participatory planning and budgeting processes, only recently has a body of evidence started to coalesce around the specific outcomes of such programming and their magnitudes.

EVALUATION PURPOSE

This evaluation will contribute to an emerging body of evidence in order to help USAID better tailor subsequent democracy and governance investments in Jordan and elsewhere. The findings from this evaluation and the learning agenda will help USAID, the Government of Jordan, and the broader donor and governance community better understand the impact of decentralization efforts at the municipal institution level and for the citizens who ultimately benefit from efforts to make government more effective, equitable, and responsive.

The primary goals of this baseline report are to explore the internal validity and baseline equivalence across institutional measures, and provide valuable context for the broader performance evaluation questions where possible. The role of this baseline report is not to provide answers to the evaluation questions, but to generate findings for comparison with later data collection rounds. The broader performance evaluation questions will be addressed in full in the endline evaluation report.

REPORT STRUCTURE

This document explores the pre-treatment data collected for the CITIES evaluation and assesses the severity of validity threats to impact evaluation measures posed by pre-treatment differences across treatment and comparison municipalities. First, it describes the Jordanian governance context and the main CITIES activity components, which have been designed to address key decentralization and local governance efforts specific to Jordan. The document then lays out the evaluation questions, theory of change, analytical approach, and key indicators for the learning agenda. The report then presents descriptive statistics and summary of qualitative information to identify any substantive pre-treatment differences that may threaten the validity of impact measurements at endline. Finally, a discussion session will investigate any substantive differences and make final assessments as to whether the validity of the evaluation design still holds.

BACKGROUND

Strengthened democratic accountability is critical to Jordan's stability and future prosperity. Over the past 10 years, Jordan has pursued several structural reforms to enhance the conditions for liberalization and more inclusive growth. Decentralization is an integral part of the reform agenda, undertaken to encourage a greater proportion of citizens to engage politically and to mitigate the stressors that are affecting them on a daily basis. Adverse regional developments, in particular the crises in Syria and Iraq, continue to affect Jordan in the form of increased socio-economic pressures due to the vast inflow of refugees.

There is a long tradition of official municipal organization in Jordan, yet developing relevant, responsive, and representative local institutions remains an ongoing challenge¹. Municipalities as a modern governance institution were developed in 1955 and have been subject to several reform efforts (Ababsa, 2003). In 2002, the Ministry of Municipal Affairs (MoMA) consolidated the number of municipalities across the Kingdom to address financial insolvency issues, with the total number of municipalities falling from 328 to

¹ For an excellent backgrounder on the historical context of decentralization in Jordan, see Jordan - Third Tourism Development Project: Annex B.

99². There are now 100 municipalities, which receive funding from the national government based in part on their population, and within which there is a municipal council voted into office through local elections.³ Municipalities collectively cover an estimated 83 percent of Jordan's population. This local administrative structure overlaps with other institutions that exist at the sub-national-level, such as governorates, and Ministry offices and initiatives, such as Local Development Units, which are responsible for socioeconomic development in municipalities and governorates.⁴

In 2015, Jordan passed both a Decentralization Law and a Municipalities Law as part of an effort to integrate municipalities into a general governance framework and increase accountability among local democratic institutions. Amman has been the locus of control and remains the primary service provider across Jordan. Recent local elections that took place in August 2017 under the new legal framework provide a window of opportunity to strengthen local governance and engage communities in their own development. The elections brought elected local councils, mayors, municipal councils and governorate councils under a new system of governance that is expected to strengthen accountability and transparency, improve service delivery and increase community engagement.⁵

The new laws for decentralization and municipalities outline three levels of local representation: the local council, which has at least five members and represent more than 3,000 constituents; the municipal council, which is formed by the heads of the local councils, in addition to an elected mayor; and a governorate council consisting of 85 percent directly-elected representatives and 15 percent appointed members. It remains to be seen if these recent legislative measures, and the increased enfranchisement at the municipal level lead to improved local governance and increased citizen engagement.

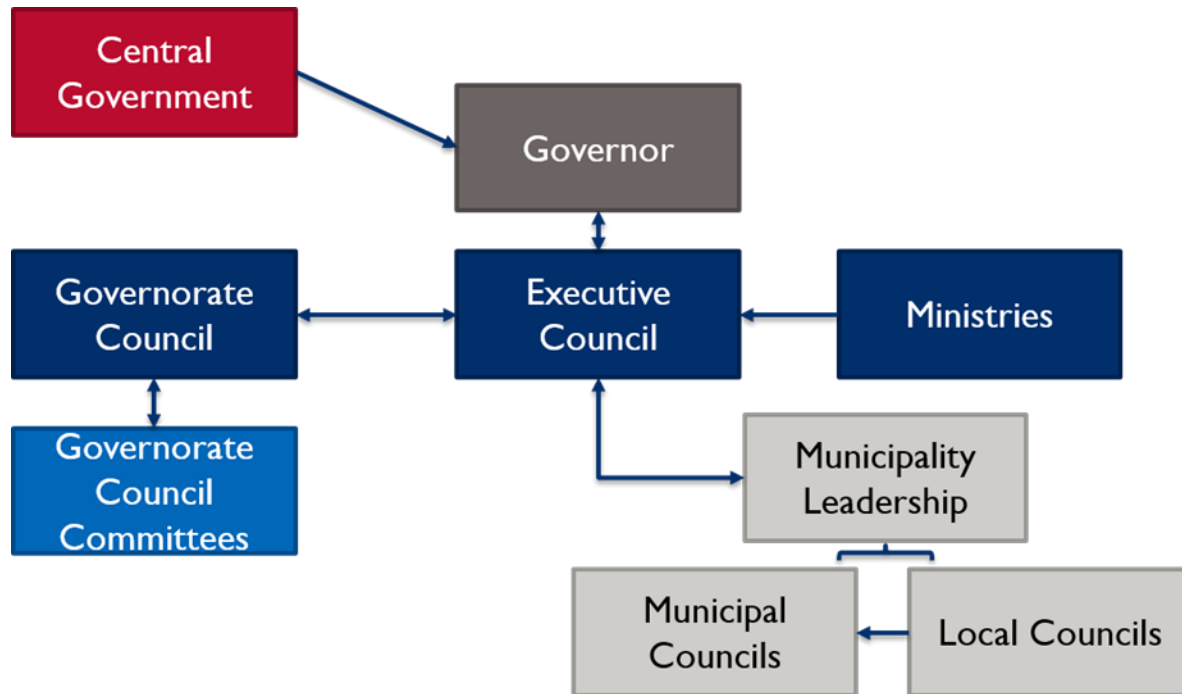
2 See Bergh, 2016, page 81. Also, MOPIC (Ministry of Planning and International Cooperation), 2006, Municipal Joint Service Councils Assessment Study, ERM, 154 p. as quoted in Ababsa 2013.

3 See Bergh 2016. Also, Jordanian Embassy, About Local Government.

4 See OECD, 2017.

5 Ibid

FIGURE I LOCAL GOVERNMENT STRUCTURE



CITIES ACTIVITY OVERVIEW

CITIES is a five-year activity implemented by a consortium led by Chemonics International with an estimated budget of \$58,549,993 and period of performance from September 25, 2016 – September 24, 2021⁶. CITIES aims to increase the effectiveness of municipal governance and support decentralization in Jordan and is implemented in 33 municipalities across all 12 governorates in Jordan. By strengthening municipal institutional capacity, improving public service delivery, enhancing municipal-community communications and strengthening community resilience and cohesion, it is expected that citizens will ultimately judge their local municipality to be more responsive to their needs and more effective in executing their core functions.

CITIES implementation is based around four components that cut across its overarching goal of supporting decentralization and increasing the effectiveness of local government in Jordan:

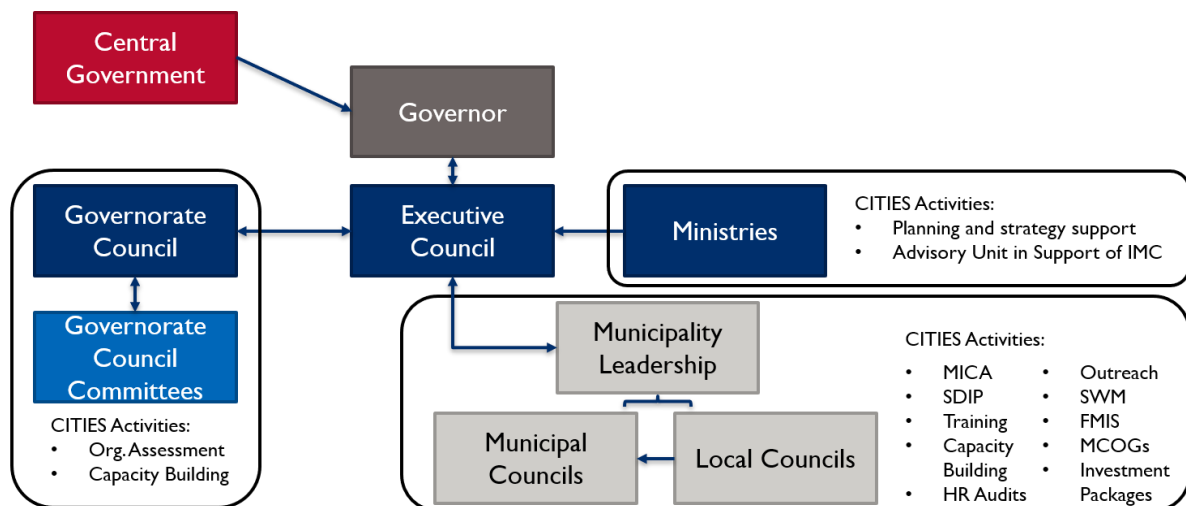
1. **Improved Service Delivery:** Illustrative activities include the Development of Service Delivery Improvement Plans (SDIP); assistance with solid waste management (SWM); conducting municipal institutional capacity assessments (MICA); and additional activities that align with MoMA priorities.
2. **Increased Effectiveness of Local Government Operations:** Illustrative activities include analysis on legal and regulatory frameworks; review of human resources, financial, organizational, and budgetary management in municipalities.

⁶ Chemonics leads a consortium of implementing partners consisting of VNG International, Planet Partnerships, ConsultUs, and Interdisciplinary Research Consultants.

3. **Increased Responsiveness to Community Priorities:** Illustrative activities include support for the Ministry of Municipal Affairs (MoMA); and assessment of public outreach tools.
4. **Enhanced Capacity to Promote Community Cohesion and Resilience:** Illustrative activities include youth engagement, and community interventions addressing Threats to Community Cohesion (TTCCs) .

By strengthening municipal institutional capacity, improving public service delivery, enhancing municipal-community communications and strengthening community resilience and cohesion, it is expected that citizens will ultimately judge their local municipality to be more responsive to their needs and more effective in executing their core functions. As shown in the figure below, most CITIES activities are focused at the municipality level of government, but implementation for component 2 and support by the crosscutting Advisory Unit for the Inter-Ministerial Committee for Decentralization (IMC) do include interventions at higher levels, as well. The results of CITIES are expected to contribute to USAID's development objective of stronger democratic accountability and a more stable, prosperous Jordan.

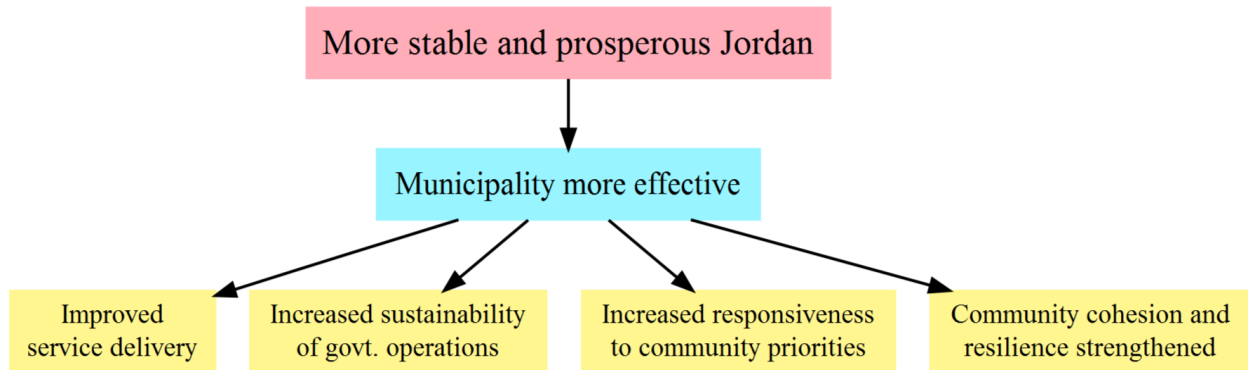
FIGURE 2 JORDAN'S GOVERNANCE STRUCTURE AND CITIES ACTIVITIES



THEORY OF CHANGE

According to the CITIES Monitoring, Evaluation, and Learning (MEL) plan, if municipal government effectiveness and service delivery is increased, and if municipalities are better able to address the priorities of their community members while promoting community resilience, then democratic governance will be enhanced; trust and confidence in elected officials will be improved; and stability will be strengthened. A simplified theory of change (with arrows denoting the 'how' rather than 'so what' question) is as follows:

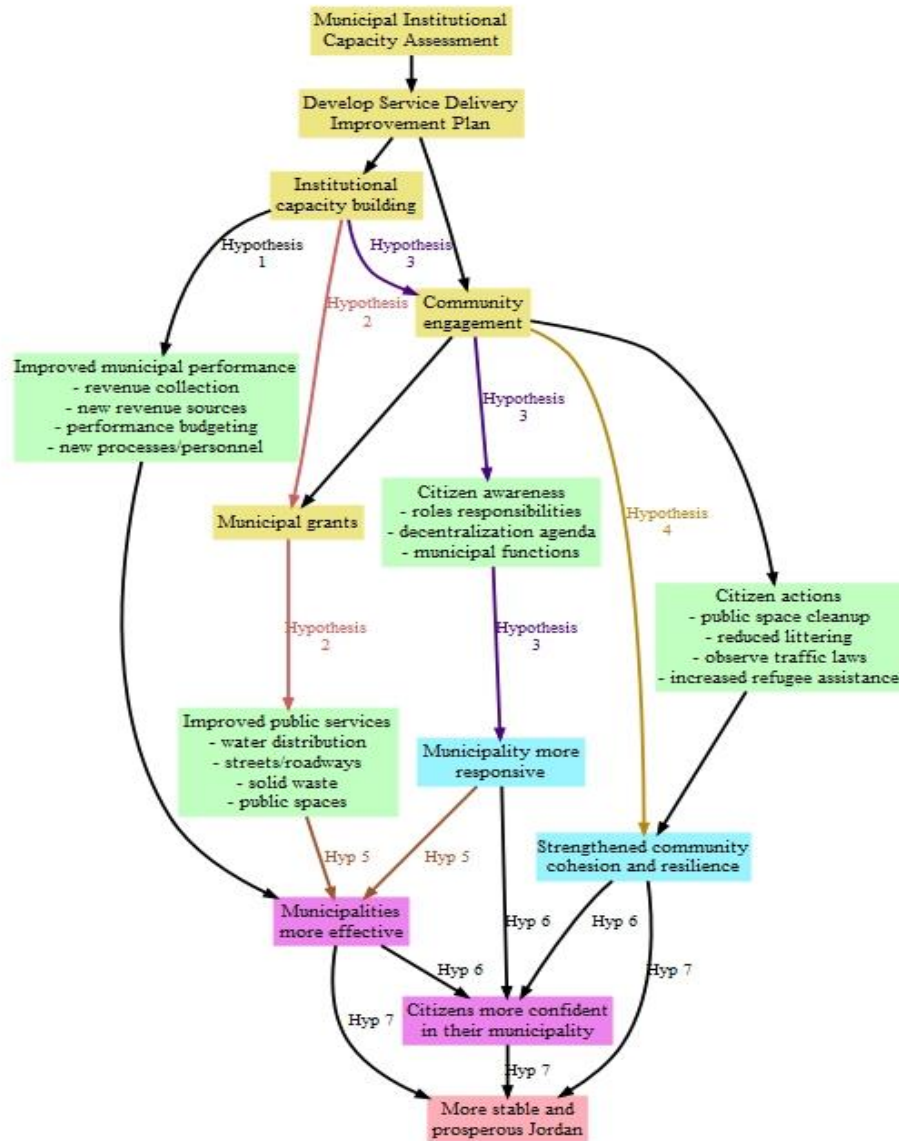
FIGURE 3: SIMPLIFIED CITIES THEORY OF CHANGE DIAGRAM



In order to more fully delineate the primary pathways by which the CITIES intervention could meet its objectives, the evaluation team expanded the logic model to help guide the evaluation. The figure below follows the various programmatic streams under CITIES and relates these through seven hypotheses to key outcomes and overall goal.

In this extended logic model, boxes in yellow represent programmatic inputs, light green boxes represent intermediate outcomes, light blue boxes represent activity outcomes, and purple boxes represent mission-level outcomes. All boxes feed into the overall USAID mission goal of a stable and prosperous Jordan. Note that the highest-level impact indicator in the CITIES MEL plan is municipal government effectiveness, while the CITIES evaluation includes the higher-level impact measures of confidence and stability. For accountability purposes, CITIES performance targeting applies only to its sub-purpose outcomes of service delivery, institutional capacity, responsiveness, and cohesion, and its activity level impact measure of municipal government effectiveness.

FIGURE 4: EVALUATION THEORY OF CHANGE



EVALUATION QUESTIONS

The ET set out to investigate the overall impact of CITIES and has also organized its work according to broad performance evaluation questions that examine activity effectiveness, sustainability, synergies, and learning. The evaluation questions are summarized below in Table I.

TABLE 1: CITIES EVALUATION QUESTIONS

Theme	Question(s)
Effectiveness	What is the overall effectiveness of CITIES for achieving its goals of enhancing democratic governance, improving trust and confidence in elected officials, and strengthening stability? Did the activity achieve its intended outcomes? Are there certain areas/activities and components that have been more effective than other areas/ activities and components?
Sustainability	Which interventions under CITIES are most likely to be sustained over time? Why and how? To what extent are CITIES capacity building efforts being sustained by municipalities and other targeted bodies of decentralization?
Synergies	How well has CITIES developed synergies (coordination and collaboration) with other USAID activities that focus on similar objectives, for example, decentralization and working with Municipalities? Are donors operating in the decentralization and local governance space achieving synergies through coordination or knowledge sharing?
Learning ⁷	Based on the performance of CITIES, both in terms of effectiveness and sustainability, what are some key lessons learned, by component and sub-activity?

To enhance measurement and learning purposes, the overall CITIES theory of change was broken down into seven constituent hypotheses based on how programmatic inputs and activity outcomes may interact to produce desired results. This also enables the ET to better assess the different aspects of activity implementation and their respective contributions to observed outcomes. However, while the validity and directionality of each hypothesis will be explicitly tested, and the evaluation learning agenda will examine how the failure of one hypothesis may affect downstream hypotheses, the figure above cannot pretend to fully capture the realities of activity implementation. Rather, this extended logic model is designed to identify hypotheses that contribute to mission learning about what works and what doesn't, and also recognizes that actual interactions from implementation will likely remain too complex to be fully diagrammed. These constituent hypotheses referenced in the figure above are presented in the table below.

TABLE 2: CITIES EVALUATION HYPOTHESES

Evaluation Question	Hypothesis
I. Effectiveness	Hypothesis 1: If CITIES provides institutional capacity building, then municipal operations will improve.
I. Effectiveness	Hypothesis 2: If CITIES provides institutional capacity building, combined with incentive grants and improved public service delivery, then this will lead to an improvement in the quality of public services.

⁷ For the baseline assessment, questions 1-4 will be the primary focus. Evaluation question 4 will be addressed at midline and endline.

Evaluation Question	Hypothesis
1. Effectiveness 2. Sustainability	Hypothesis 3: If CITIES provides institutional capacity building, combined with proactive outreach by municipalities, then this will increase public perception that municipalities are responsive to citizen needs.
1. Effectiveness	Hypothesis 4: If CITIES supports increased attention to marginalized groups, such as youth, gender, nationality, and disability, then this will strengthen a sense of community cohesion and resilience.
1. Effectiveness 3. Synergies	Hypothesis 5: If CITIES provides institutional capacity building, combined with proactive outreach and increased / improved public service delivery, then this will increase public perception that municipalities are effective.
1. Effectiveness 2. Sustainability 3. Synergies	Hypothesis 6: More effective public service delivery, responsiveness to citizen needs, and strengthened community cohesion and resilience will increase citizen confidence in their municipality.
1. Effectiveness 2. Sustainability 3. Synergies	Hypothesis 7: More effective public service delivery, confidence in local leadership and performance, and strengthened community cohesion and resilience will lead to a more stable and prosperous Jordan.

EVALUATION DESIGN

As described above, this evaluation is a mixed method quasi-experimental design that focuses on both impact and performance aspects. Key aspects of the evaluation design include:

1. Evaluation Indicators
2. Matching process for the construction of a comparison group using contextual indicators
3. Quantitative Data Collection and Analysis
4. Qualitative Data Collection and Analysis on the state of local governance and decentralization
5. General Population Survey
6. Implementation Fidelity
7. Learning Agenda

This section provides a brief overview of these design features.

EVALUATION INDICATORS

Across all aspects of the evaluation, a set of 10 indicators was designed to measure each hypothesis outlined in the evaluation team's interpretation of CITIES' theory of change. Each indicator has been mapped to the evaluation questions (presented here by their themes for brevity), the data source, disaggregation approach, and analysis method. The indicators are as follows:

TABLE 3: EVALUATION INDICATORS

Hypothesis	Indicators	Data Source	Disaggregation
Hypothesis 1: Institutional capacity building will improve municipal operations.	MICA Score Services MICA GESI Score Municipality Revenue Budget allocation to salaries ⁸	• MICA assessment data • Available municipality administrative data • KIs	Municipality, district, governorate, region
Hypothesis 2: Institutional capacity building, combined with incentive grants, will improve the quality of public services.	Availability of services Perceived quality/availability of services	• Beneficiary survey • Population survey • KIs • FGDs	Service, municipality, district, governorate, region, gender, age, education
Hypothesis 3: Institutional capacity building, combined with proactive outreach by municipalities, will increase public perception that municipalities are responsive to citizen needs.	Perceived municipal responsiveness ⁹	• MICA assessment data • Beneficiary survey • Population survey • FGDs • KIs	Municipality, district, governorate, region, gender, age, education
Hypothesis 4: Increased attention to marginalized groups such as youth, gender, nationality, and disability will strengthen a sense of community cohesion and resilience.	Perceived municipal effectiveness ¹⁰	• Beneficiary survey • Population survey • FGDs • KIs	Municipality, district, governorate, region, gender, age, education

8 Note: These data were collected as part of the MICA data collection, but the evaluation team found significant data quality issues with the self-reported budgetary and revenue data. Future rounds of data collection will utilize statistics from MOMA.

9 Municipal responsiveness can be defined as municipality governance listening to and addressing peoples' needs in an open and efficient manner.

10 Municipal effectiveness can be defined by perceptions of the degree to which municipal leaders and administrators can formulate and implement policies that address citizen needs, the perceived quality and maintenance of existing and requested public services, and the perceived credibility of the government to follow through on citizen needs.

Hypothesis	Indicators	Data Source	Disaggregation
Hypothesis 5: Institutional capacity building, combined with proactive outreach and increased / improved public service delivery, will increase public perception that municipalities are effective.	Perceived ability to address common problems ¹¹	• Beneficiary survey • Population survey • FGDs • KIIs • Activity and municipality documents	Municipality, district, governorate, region, gender, age, education
Hypothesis 6: Improved public service delivery, increased public perception of municipal responsiveness and effectiveness, and a stronger sense of community cohesion and resilience will lead to an increase in public confidence in municipal / local governance.	Confidence in local government Trust in local government	• Beneficiary survey • Population survey • FGDs • KIIs • Other activity documents	Municipality, district, governorate, region, gender, age, education
Hypothesis 7: More effective public service delivery, confidence in local leadership and performance, and strengthened community cohesion and resilience will lead to a more stable and prosperous Jordan.	Quality of life Composite measures	• Beneficiary survey • Population survey • FGDs • KIIs • Other activity documents	Municipality, district, governorate, region, gender, age, education

MATCHING FOR COMPARISON GROUP CONSTRUCTION

In the absence of randomizing the assignment to treatment, grouped data by beneficiaries and non-beneficiaries is not directly comparable due to observable or unobservable differences between groups in addition to treatment status. One wishes to know how the treatment affects beneficiaries but cannot be sure if any observed group differences are driven by treatment or some other observable or unobservable difference between groups.

A common tactic in such a situation is to use observable characteristics to attempt to adjust for these differences and thereby generate “as-if” random assignment allowing a direct comparison of the outcomes

¹¹ The ability to address common problems refers to the perception of how political, tribal, gender, religious, or other affiliations influence a municipality’s response to citizen needs and requests, as well as the ability of municipality leadership to coordinate without inequitable concession to these affiliations.

between the two groups as the estimate for activity impact. Under this approach, researchers create pairs between each implementation unit and a comparison unit on the exact same observed characteristics; however, this can quickly become onerous and difficult if the number of observable characteristics required to find one-to-one matches is large. One alternative method of applying this tactic is to select a sub-sample of your data beforehand using Propensity Score Matching (PSM). Under Propensity Score Matching, a select number of observable characteristics are used to generate a predicted probability of belonging to the treatment group. This predicted probability (the propensity score) is then used to match treated and comparison units based on the theorem that if they match on the score generated from observable characteristics, they essentially match on the specific values of the observable characteristics as well.

CITIES is being implemented in 33 municipalities, including all 10 of Jordan's Class A municipalities. This left 23 municipalities available to match. Time and cost constraints limited the potential pool of comparison municipalities to 15, and matching was done on the following observable characteristics:

- Poverty rate
- World Bank municipal grants (ESSRP)
- Refugee population
- North region
- Community Engagement Project
- Class C
- Log population
- Class B
- South region
- Central region

The matching analysis described in the design proposal suggests that on these observable metrics, there is statistical equivalence between 15 CITIES municipalities and 15 comparison municipalities. As discussed later in this report, it is important to re-visit the assumption of equivalence on the outcome of interest at baseline to ensure this similarity holds and to determine what adjustments may be needed at endline. The design report provides a more detailed summary of the matching process.

QUANTITATIVE DATA COLLECTION & ANALYSIS APPROACH

The primary tool utilized to collect data to be analyzed across the two assignment groups is the Municipal Institutional Capacity Assessment (MICA) tool. Chemonics, the CITIES implementing partner, originally developed a version of the MICA as part of an activity in Afghanistan, and then adapted the tool for the Jordanian context.¹² The MICA, according to CITIES staff, was utilized as a way to “start a conversation” around its implementation approach in each of its municipalities, and was only a first step to build a picture of where services can be improved.¹³ Overall, MICA implementation involves collecting information on

¹² It is important to note that the MICA tool was not rigorously pre-tested prior to implementation in Jordan. Adaptations to the tool could be expected based on learning from each round of data collection by CITIES and/or the evaluation team.

¹³ After scoring the municipality, CITIES staff facilitates the development of the Service Delivery Improvement Plan (SDIP), which is a demand-side process that helps community members identify priority interventions for CITIES to support. The SDIPs will incorporate issue areas identified through the MICA and will also reflect other stakeholder concerns not captured in the MICA. Note, further, that other CITIES components have their own assessment methods, ranging from financial management assessments that closely mirror MICA in terms of systematically scoring competencies, to more rapid qualitative assessments.

each MICA element through interviews with key municipality staff (e.g. Executive Director, and others), verifying information to the extent possible by viewing relevant municipality documentation, and assigning a score for each MICA element, based on a scoring rubric. From this tool, element specific and aggregate scores are generated for each municipality. Summary notes justifying each score are included in any final scoring document.

Change in the MICA scores then serves as the empirical measure used to document changes at the municipal institutional level.¹⁴ For the ET, MICA scoring items also guided some of the qualitative inquiry to better understand how the context and in relation to the performance evaluation questions. Given the importance of MICA scoring for both implementation and estimation of CITIES's impact, assessment of the MICA scores themselves is warranted to ensure reliable, minimally biased estimates.

As detailed in the design report, the basic measurement to estimate outcomes between the two assignment groups will follow a difference-in-differences (d-i-d) approach, in which a sample of both treated and untreated units (in this case, municipalities) are examined before (baseline) and after (endline) treatment. Examining the first differences in outcomes across both groups (trends) rather than at single points in time (levels) has the important effect of removing any unobserved sources of bias in the data that are common to both groups and remain constant over time. The difference between the trend lines of each group (the second difference) then generates an estimate of the overall treatment effect. Adding additional terms to distinguish between direct and indirect beneficiaries will extend the analysis to examine the extent to which CITIES activities are visible to municipal populations.

To better understand the implementation context beyond quantitative metrics, this evaluation also features a significant qualitative portion. While the impact evaluation is focused at the municipality level, systemic processes and bottlenecks at higher levels of governance are expected to affect municipal effectiveness and general performance. This is due to potential overlap between the deconcentrated structures and the lack of regulations clearly delineating their roles and responsibilities. For example, under the new deconcentrated system, municipal capital needs will be identified through a bottom-up process but will be harmonized/reconciled at the governorate level depending on factors that are not yet identified. These factors, associated regulations, and political economy considerations will determine what needs are eventually met at the municipal level. In turn, the municipal “needs” that are eventually funded will have a direct bearing on how well the community perceives the municipality and the quality of its services. CITIES expected outcomes will therefore be affected by external factors beyond the locus of the Activity's interventions. While some of these factors may be picked up through the quantitative measures, such as select MICA items, measurement may be more akin to a proxy rather than a direct record of high-level changes. Moreover, qualitative tracking is needed to understand and map the technical, regulatory, and relational dimensions of the various municipal functions mandated in the Municipalities Law. Tracking decentralization reforms and collecting contextual data will be important to understanding the municipalities and CITIES' ability to improve these functions.

The ET analyzed data from KIIs to highlight key themes of interest across the interviewees and discussion participants. The team then summarized responses related to each theme and included quotations from respondents to illustrate key findings. Beyond the baseline, the team will also make planned/actual

¹⁴ Other measures, such as citizen perception of local government responsiveness and effectiveness, are more impactful. However, improvement in MICA scores is directly within CITIES' manageable interests to achieve, while changes in citizen perception are subject to multiple factors outside of CITIES' control and are also difficult to measure with the necessary precision to attribute to the CITIES activity.

comparisons in answering evaluation question one. The ET will make comparisons between activity descriptions, work plans, targets, and periodic performance data to inform the examination of CITIES's performance relative to overall perception of the activities.

GENERAL POPULATION SURVEY

One of the key data sources for this evaluation is a large-scale survey that USAID/Jordan commissioned to collect mission project and activity performance data, as well as contextual data, from May-September 2018. The survey measures cover six mission performance indicators across multiple sectors, including CITIES-related service delivery and citizen engagement items measuring Hypotheses 2-7. The ET will draw from this general population survey to estimate household-level changes in both CITIES and non-CITIES municipalities to test hypotheses on citizen engagement and government responsiveness and effectiveness.

For purposes of this baseline report, the CITIES-relevant measures are assessed for their balance properties across treatment and comparison municipalities, general awareness of and knowledge about decentralization, and information about how active and engaged citizens are with local government or broader policy issues. See the section below on public perception indicators as well as relevant annexes for presentation of population survey information.

IMPLEMENTATION FIDELITY

Ensuring that the basic assumptions about implementation hold is critical to the internal validity of any evaluation design. The ET has undertaken regular implementation fidelity checks to ensure that any potential risks to the validity of the design are discovered quickly and any changes in implementation, from actual activities and scope to the general timeline of CITIES, are discussed early enough to address in a methodologically sound manner.

The risks of ignoring implementation fidelity range from potentially misunderstanding aspects of activities to increasing error in estimates and reducing the effect size that endline analysis detects. To avoid these issues, the ET requested access to the regular reporting of CITIES (e.g. quarterly and annual reports). The ET continues to work with USAID and CITIES for check-ins with the CITIES staff to address any potential challenges or changes based on review of the quarterly reports and follow-up on developments in the municipalities.

TABLE 4: IMPLEMENTATION FIDELITY PLAN

Activity	Description	Status
Status Updates	Formal discussion of implementation and evaluation progress, highlight changes, discuss potential next steps or updates.	Sessions conducted in Feb and Aug 2018
Implementation Tracking	Work with MEL team to develop tracking inputs. This will be a one-off meeting initially and then incorporated into the quarterly status updates.	In progress

Activity	Description	Status
Document Distribution	Sharing of implementation documentation, such as municipality reports and work plans.	In progress
MICA synthesis	Shared review of MICA scores between implementation and comparison areas to ensure consistent implementation of the tool and to walk-through updates or changes enacted by CITIES	Conducted Aug 2018
MEL Data check	Review of municipal rosters, Component 2 HR audits, Component 1 MICA data, other MEL data. Share data from surveys, activity monitoring, municipality tracking, and other activities; the evaluation team will then input this data as part of their regular tracking for analysis	In progress. CITIES baseline data shared in Feb 2018, municipal roster data shared in Aug 2018
Fidelity Interviews	Brief KIs with municipality administration and leaders to understand implementation adherence, delivery, and differentiation.	To be completed as part of midline evaluation (~ Oct 2019)

The goal of implementation fidelity is not to evaluate CITIES implementation through some additional approach or make planned versus actual judgments, but rather to ensure that if and when changes to implementation occur, the ET can adequately and accurately account for them in their analysis and design. Any qualitative fidelity interviews with municipality staff will follow a semi-structured format based on a general interview template to be developed based on review of the CITIES work plans. The quarterly check-in calls with CITIES will be unstructured and allow both the evaluation team and CITIES staff to openly discuss potential challenges and progress on both sides.

LEARNING AGENDA

In addition to the above approaches, the evaluation also includes an exploratory learning agenda to help understand the implementation environment in which findings are situated, identify heterogeneous treatment effects, and learn which factors strengthen or impede activity success. A set of learning agenda questions were developed and can be referenced in the Evaluation Design. Ultimately, these questions will inform what can be learned analytically through a triangulation of data sources and secondary data analysis. Additional lines of inquiry may be added as the evaluation team learns more about implementation, and CITIES adapts to local municipality needs. This analysis is anticipated to take place primarily at midline and endline. Initial findings from the learning agenda analysis using the general population survey can be found in Annex 9.

EVALUATION LIMITATIONS

As outlined in the evaluation design proposal, there are several key limitations to the approach of this evaluation.

- **Unobservable factors and data scope:** the quantitative portion of this evaluation relies on the MICA tool to determine changes in municipality management and service provision. This tool was designed to help CITIES target its implementation, and not to fully capture municipality performance. There may also be factors that affect MICA scores that are not part of CITIES's implementation, and there may be municipal capacity development that is not captured in the MICA. For example, social connections may affect item-level changes on the MICA but would not be captured. The ET's qualitative data collection aims to fill the gap, but still relies on respondents being aware of or forthcoming about otherwise unobservable or indirect factors. In addition, the effect of any given MICA item on citizen perception is unclear. Aggregate scores are the main outcome of this evaluation, but it remains to be seen through qualitative data how item-level and aggregate MICA scores relate to perceptions and experiences.
- **Nature of the MICA Tool:** The MICA was designed as both a rapid assessment tool and a baseline measure of municipal performance whose improvement over time is a performance indicator in the CITIES MEL plan. This leads to a natural tension between the immediate impressions of a rapid assessment and a robust measurement of capacity that would be expected of an impact estimate. This natural tension should be kept in mind when assessing endline estimates, at which time the ET should expect to gather valuable lessons learned about the efficacy of such an approach in evaluation.
- **Internal validity:** The MICA items were adapted from uses in several countries but not rigorously pre-tested in the Jordanian context prior to use, leading in a few cases to issues with how items are framed or scored that would most probably be corrected in any future version of the instrument. There is also a degree of scoring interpretation left to the scorer, creating an inherent subjectivity that adds to uncertainty around measurements that could magnify across different scoring teams such as the CITIES scorers and the MESP scorers. These are non-trivial risks to the internal validity of MICA measures which necessitated extra care. The ET held multiple trainings with CITIES, applied multiple raters to independently score each comparison municipality, and reviewed scores with the CITIES Component I team lead. However, while initial indications are that the MICA scores are of sufficient quality to use in the evaluation, the internal validity of the data will be critical to monitor throughout the evaluation's data collection rounds.
- **Recency and recall biases:** The ET will utilize administrative and implementation data, such as MICA scores, for its estimation of CITIES impact. However, a large portion of understanding how decentralization is progressing involves asking people about their perceptions of how decentralization is progressing. This creates several challenges. Notably recency bias, wherein people remember whatever has happened most recently, whether good or bad, rather than what has occurred over a certain period of time. Recency bias can be an issue even when questions are asked about, "the past 12 months." Similarly, recall bias relates to the fact that people do not always remember what activities have taken place or the details of their involvement. For example, people may not recall that municipal leaders held outreach events or introduced new street cleaning activities if they did not occur in the proximate past. The use of multiple data sources, as well as multiple data collection periods, will help mitigate some of the effects of these biases. The use of well tested survey instruments can also help improve responses and diminish the potential risk these biases pose.
- **External validity:** The nature of the CITIES activity imposes limitations on the external validity of the evaluation (i.e., the extent to which the evaluation findings can be expected to

predict the impacts of a program like CITIES undertaken in a different country or in Jordan at some point in the future). First, as noted in the background section, Jordan has undergone multiple institutional reform efforts and it is likely that changes to the local governance context will continue to evolve. The results from this IE may not hold in a future context or even at a different governing-level. Secondly, any results from an IE come with critical caveats that often get forgotten in the shadow of key findings. The ET will strive to make the limits and exceptions of their analysis clear. Causal attribution can be highly sensitive to underlying assumptions and analytical specification. The extrapolation of the IE results to a broader understanding of decentralization efforts may not be valid and should be informed by detailed understanding of this report.

- **Changing implementation context:** The ET has worked closely with CITIES to track implementation and understand the implications for this evaluation. In some cases, this has meant noting when activities are applied across both assignment groups; in others, it will mean additional qualitative inquiry during future data collection rounds to see how perceptions of need have changed over time and if this has led or lagged implementation activities or broader changes in Jordan. CITIES implementation approach is responsive to local needs, which is valuable for beneficiaries as it likely increases the relevance of activities. If interventions are provided that, for example, do not directly affect MICA scores it may be the case that no impact is found when positive changes did in fact occur; as noted above, the MICA may not measure all of CITIES activities. Strong communication between implementation and ETs and implementation fidelity tracking will help mitigate this limitation, but it still may be the case that changes in CITIES approach occur that are not controlled or accounted for through the ET's analysis.
- **Failure to capture a true baseline:** In a few municipalities, CITIES initiated activities prior to the MICA measurements in comparison municipalities.¹⁵ For all CITIES municipalities, the public perception survey was conducted after CITIES initiated its work but prior to execution of grants that would be most likely to affect public perception. While these violations to the evaluation design are not considered significant validity threats, the ET must be alert to the possibility of substantive differences between treatment and comparison municipalities being caused by the onset of CITIES activities. This most typically will have the effect of underestimating any treatment effects as there will be some effect of programming that is captured as a baseline difference rather than a treatment effect.
- **Other donor activity:** The ET estimates that there are currently 28 implementing partners operating across 48 activities in the democracy and governance space in Jordan. The evaluation will be able to control for the biggest actors such as CEP, ESSRP/MSSRP, but will not be able to track all donor activity that may also affect CITIES outcomes. If untracked donor activity realizes substantive outcomes, it will distort the CITIES impact estimates in ways that depend on the distribution of that donor activity. For example, if there are untracked outcomes realized in comparison municipalities (the most likely scenario), it will have the effect of a) underestimating the CITIES treatment effect under the assumption of comparing CITIES to no intervention, or b) will shift the impact estimate to be how much

¹⁵ The primary activities conducted prior to the start of this assessment include: Regional Youth Awareness Sessions (Greater Amman, Wadi Mousa, and Greater Irbid), Decentralization and Community Engagement Awareness Sessions (12 governorates), and a training for LDU staff on Youth Engagement and Development in Al Hallabat municipality. Other efforts by CITIES prior to the implementation of this baseline study focused on various assessments, including MICA and SDIP development for treatment municipalities.

CITIES is to the alternative of other donor programming. The safest interpretation under this threat is to assume that the comparison municipalities have some level of other donor programming relevant to CITIES outcomes.

BASELINE DATA COLLECTION

To measure across evaluation indicators, MESP collected data utilizing both quantitative and qualitative data collection as part of the baseline. A summary table is included below.

TABLE 5 BASELINE DATA COLLECTION METHODS, TIMELINE, AND INDICATORS

Data collection activity	Scope	Timeframe	Indicator(s)
Municipal assessments (MICA)	33 CITIES treatment municipalities MICA scores collected by CITIES	Treatment Municipalities: April 2017-January 2018	•MICA Services score
	15 comparison municipalities MICA scores collected by the evaluation team available administrative data for all municipalities	Comparison Municipalities: February – March, 2018	•MICA Gender and Social Inclusion (GESI) score
Key Informant Interviews	20 KIIs conducted with government (national, local), governorate and local councils, donors, local CSO, and CITIES	March-June, 2018	•Baseline contextual information (political economy, decentralization, MICA services areas, donor activity)
Population survey	Municipality population, partitioned by direct and indirect beneficiaries within treated municipalities (June - September, 2018)	June – September, 2018	•Quality / extent / satisfaction with govt. service, disaggregated by specific service and passive / direct users
			•Awareness of development activities / local govt. outreach in their municipality •Perceptions of local governance •Knowledge / awareness of local governance

QUANTITATIVE DATA COLLECTION

As mentioned above, the MICA is CITIES's primary tool for determining a municipality's service delivery capacity and a key quantitative input for the evaluation within its target municipalities (treatment municipalities). CITIES provided MICA scores for each of their municipalities to serve as the treatment

group. Building upon the efforts of CITIES, the ET used the same CITIES MICA tool to collect equivalent quantitative data in each of the 15 comparison municipalities. For more information on the scoring process, see the inset below, and Annex 2.

The ET used the CITIES MICA tool to score each comparison municipality using an ODK form. The ET used the same form to re-enter MICA scores from CITIES's municipalities, as well, to create a data set that contains the full universe of all MICA scores, including those municipalities that were not identified as strong matches through the initial municipality matching process.

The final set of scores produced internally within the ET scoring of the 15 comparison municipalities was subjected to several rounds of review between enumerators, and as discussed in Internal Validity of MICA Scores section, the evaluation team conducted an analysis of its scoring across multiple MICA raters to develop a final set of comparison municipality scores. Upon arriving at final scores in comparison municipalities, the ET conducted a final review of scores and their justification with the CITIES Component I lead to ensure reliability and consistent scoring. This led to a few modifications of comparison scores in order to harmonize the interpretation of scoring criteria across the CITIES and evaluation teams. This scoring was labeled as 'Alternate' and maintained as a separate set of outcome values for the purposes of conducting sensitivity analysis. While the alternate scoring would be the preferred choice among either, ideally any CITIES treatment effect should persist across both sets of scores. The values and figures in this document report the "Alternate" scoring that was determined in coordination with CITIES.

MICA SCORING AND SCORE UNIFICATION PROCESS

The MICA tool scores municipalities from 0 to 5 across 23 items related to municipal services, and 0 to 1 on 17 items related to Gender and Social Inclusion (GESI).

The ET worked with CITIES to mimic their overall MICA implementation approach. The general MICA scoring process works as follows:

1. **MICA Training:** An item-by-item review and discussion of different scoring scenarios. The ET joined CITIES on several MICA data collection sessions to observe and learn more about the data collection process.
2. **Implementation:** Meetings with municipality leadership, including mayors, council leaders, and other relevant staff, to conduct MICA. On average, each MICA interview required approximately 2 hours to complete for the ET but may require multiple visits or follow-up. For the ET, interviews were conducted as a group (2-3 enumerators), but notes were recorded separately.
3. **Scoring:** Each scorer reviews all notes taken during implementation and initial scores are given. The scorer may discuss scores with a technical lead or peer, as well as make follow-up calls to the municipality to corroborate or clarify information. Final scores are submitted. For CITIES, the technical lead conducts a final review and any changes or clarifications are made before scores are finalized.

For the ET, additional steps were added:

4. **Synthesize scores:** Each ET enumerator initially scores each municipality independently, based on their interview notes. After review of scoring differences, the scorers discuss and align

- their scores, making any additional municipality follow-ups as needed. This results in a preliminary final score for each item in each municipality.
5. Confirm scores: to ensure alignment in implementation with CITIES use of the MICA tool, the ET reviews all their scores and notes with the CITIES Component I lead. A final set of scores are developed based on this step to ensure that both the evaluation team and CITIES have comparable scoring approaches.

This process results in three sets of scores that are combined into a single dataset for accessible analysis:

- CITIES generated scores
- Synthesized evaluation team scores
- Final confirmed scores

QUALITATIVE DATA COLLECTION

As part of the CITIES IE, qualitative data was collected in both treatment and comparison municipalities. In addition to supplementing quantitative data and addressing evaluation questions, qualitative tracking was needed to analyze the technical, regulatory, and relational dimensions of the various municipal functions mandated in the Municipalities Law. Tracking decentralization reforms and collecting contextual data was also important to understand the municipalities and CITIES' ability to improve these functions.

The ET conducted 20 individual KIs drawn from: GoJ line ministries; local government (heads of governorate councils and executive council members); municipal councils and staff; local councils, and donors who are implementing projects in the same programmatic space as CITIES. Interviewees were selected based on their affiliation and familiarity with the CITIES' focus areas and ability to address questions related to the project's interventions and inform learning.

Respondents included a cross-section of officials at the national, municipal, and local levels. KIs with central and local government representatives, municipal officials and donors included the following:

- Mayors of Zarqa, Manshiyyet Bani Hassan, Mazar Al-Shamali and Al-Khaldiyyah municipalities;
- Municipal council members from Zarqa and Al-Jeeza municipalities;
- Local council members from Deir Alla and Al-Jeeza municipalities;
- Heads of the governorate councils of Zarqa and Al-Tafileh governorates;
- Members of the executive councils of Irbid and Madaba;
- Six government officials from MoMA;
- Local CBOs in Al-Jeeza and Al-Tafileh municipalities;
- Donors including the Federation of Canadian Municipalities (FCM), the Organization of Economic Cooperation and Development (OECD), World Bank (WB), the Spanish Agency for International Development Cooperation (AECID) and the French Development Agency (AFD).

The instruments for qualitative data collection are presented in Annex 8. The guides were designed to preserve the potential for a relatively free-flowing conversation, while creating a standardized format to facilitate a reliable and comparative analysis of data pertaining to the evaluation questions for triangulation of information. While the questions in the guides were based on the study's overarching objectives, they were tailored depending on the identity of the informants.

KEY FINDINGS

EVALUATION QUESTION 1: EFFECTIVENESS

Baseline data suggest that factors beyond the control of municipalities have a strong effect on municipal capacity and service delivery.

Findings related to EQ1 fall into the following thematic categories relevant for understanding the municipality context and factors that may affect CITIES' overall effectiveness:

- Municipal Services and Control
- Community Engagement within Municipalities
- Role of Local Councils
- Role of Governorate Councils
- Perceptions of Decentralization

Municipal Services and Control: According to municipal staff, the municipal sector remains mired in challenges with both external and internal factors affecting its ability to provide quality services to constituents. Weak financial stability, unclear and unfunded service provision mandates, and other issues detailed below were all reported to affect municipal services and control.

Weak financial situation: The dearth of resources and overall weak financial situation are among the municipal sector's main challenges. Weak tax collection and the limited ability to raise other internal revenues; the lack of control over government appropriations as well as a significant portion of revenue diverted to salaries, electricity and sometimes extra-jurisdictional costs the central government will not bear, limits the range of activities and services municipalities can engage in. "We can only provide waste management services," a municipal council member commented. Furthermore, these factors and others below, play out in a largely dysfunctional financial system that allows a large share of salary costs in municipal budgets—48 percent and 45 percent of budget expenditure in the comparison and treatment municipalities, respectively, went toward salaries—and limits the ability to raise internal revenue.

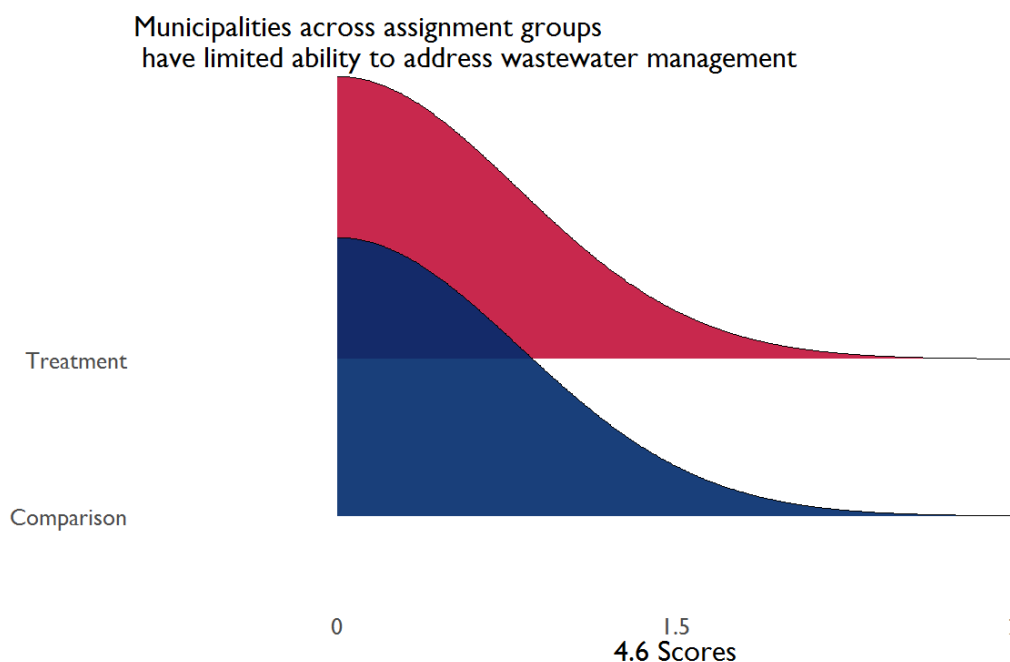
Service Provision Mandates: During qualitative interviews, several executive directors said their municipalities are providing services beyond their territorial jurisdictions because for years the government has not expanded the borders of the municipalities' regulated areas. "The biggest challenge we face is that the expansion of the regulated areas was last done in 1995. Fifty percent of the municipality is outside the regulated area. Even though we are not allowed to service these areas, we do," the head of a Local Development Unit (LDU) said. This jurisdictional issue was raised during MICA data collection, as well, further complicating how to properly score items related to service provision.

Municipalities are also often required to service main roads the Ministry of Public Works and Housing or the Electricity Company are mandated to cover. "Many ministries such as transportation, public works and water are not doing their job, especially MoMA. When they do not fulfill their responsibilities, our own performance is affected," an executive director said. A mayor subsequently commented on the effect this has on citizens' perceptions of the municipality: "The Amman-Jerash road is a problem. People think we are responsible for the street when we are not. The Ministry of Public Works and Housing is not responsive. People tell us the health center does not have medicines. This is really not our job."

In addition to the above data from qualitative interviews, the MICA tool captures data on items that are outside of the span of control for municipalities. This allows for some triangulation between the key informants who reported citizen expectations for services for which the municipalities are not responsible and third-party municipal assessment. MICA item 4.6, for example, covers wastewater management, which is under the control of the Water Authority. A score of zero suggests that municipalities do not have a sewage network, with scores one through three given based on the percentage of buildings connected to the sewage system. A score of four is given if there is a sewage system and the ability to address complaints related to the system; while a score of five is given if the sewage system is connected to a wastewater treatment plant with effluent that can be re-used. Each of these scores, from the presence of a sewage system to a treatment plant, deal with a public good over which there is limited local control or input.

As shown in the figure below, the assignment groups were almost identical for item 4.6, with scores heavily skewed toward zero. Both the treatment and comparison groups had a single municipality that received a score of two—Sahab and Manshiat Bani Hasan, respectively—but aside from these, MICA data collection confirms the perceptions among KII respondents that municipalities cannot deliver certain key services on their own despite citizen expectations.

FIGURE 5: MICA ITEM 4.6 SCORE DISTRIBUTION BY ASSIGNMENT



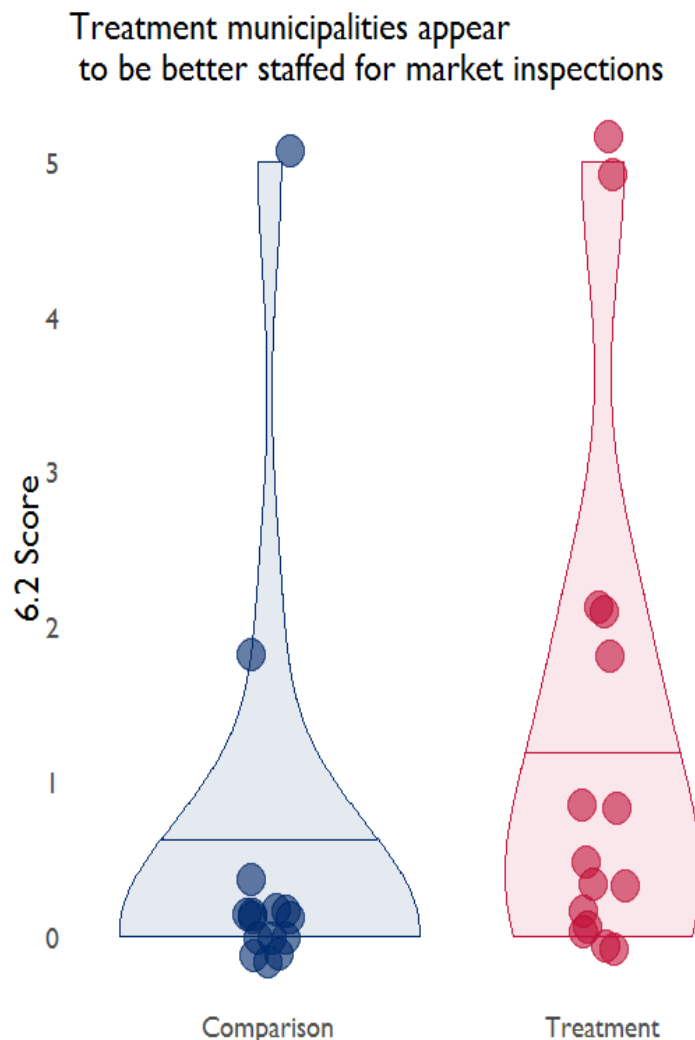
Limitations to autonomous municipal action: The lack of autonomy to make decisions on key services also limits municipal performance. An executive director said that municipalities do not have independent authority to make investment decisions. Another LDU head said that his unit was able to attract an investor and work out the parameters of a lucrative investment project, which the ministry later refused for reasons he does not understand and cannot attest to. Another executive director said: “Most of the land is agricultural. Agriculture law prohibits use of land for other purposes. When we get investment opportunities, we don’t have land for them.”

Limitations to autonomous municipal action extend to the ability to hire and fire staff. Municipal respondents said they have little say about the qualifications or selection process of new employees. “I

ask for one [staff], civil service sends me one. I have to choose that person whether qualified or not. I don't get options," one mayor said. The executive director of another municipality said: "MoMA determines qualifications for various positions. If we ask for a university graduate surveyor, they might say no. Tawjihi is enough. The health inspector in the municipality does not even have Tawjihi." Commenting on the inability to manage unwanted staff, a municipal council member commented: "We have around 4,600 employees. Many of them were appointed to appease someone. I believe half of them are at home receiving their salaries while doing nothing. This is a problem that has been going on for years."

The municipalities' functional capacity to provide quality services is circumscribed by the ability to recruit and retain municipal staff. It is also limited by the municipalities' ability to assign staff to specific positions and to determine their qualifications and job descriptions. Regardless of specific needs determined by the municipality, the central government makes final decisions as to the number and positions of new municipal staff. "We asked for a surveyor. MoMA changed the job position to a civil engineer. I already have 3 engineers. What I need is a surveyor. We know best what the municipality needs, but MoMA insists on interfering," one mayor commented.

FIGURE 6: ITEM 6.2 MICA SCORES BY ASSIGNMENT



Quantitative data show that staffing challenges differed depending on the sector or focus area. For example, MICA item 7.2 addresses whether or not qualified staff are available to respond to calamities. Six treatment municipalities received a score of zero, i.e. there are no qualified staff available for calamity management, and only one treatment municipality received a score of three, suggesting there is a calamity plan, but staff do not receive training on it. In contrast, item 6.2, which addresses whether the municipality has qualified staff available to take measures to safeguard public health and prevent diseases (hospitals, health centers, slaughterhouses, cattle markets, etc.), shows that 12 comparison municipalities did not have any staff to fill this role. As shown in the figure to the right, most of the comparison municipalities received a score of zero for item 6.2, with the average item level score of 0.5 compared to 1.3 for treatment municipalities. Other MICA items,

such as 5.1 (market inspections), also directly address staff qualifications. For MICA item 5.1, 12 comparison municipalities received a score of three, implying the municipality carries out regular market inspections and defines which goods can and cannot be traded. In contrast, only four treatment municipalities received a score of three for this item, with ten of the treatment municipalities receiving scores below three.

Most municipal respondents highlighted the staff's overall weak capacity and the need for technical support. Respondents reported the lack of technical staff to perform municipal tasks and the weak incentives to perform. Staff are regularly subjected to internal secondments that leave them holding several municipal positions at the same time. Category A municipal staff who are university graduates and tend to relatively be more motivated do not exceed 5% of the number of overall municipal staff. More problematic is that this "municipal leadership" cannot under the current system and circumstances hire and fire staff members. There are social and legal constraints that limit the ability of the higher echelon of staff to address this problem. One mayor in the north commented:

"I'll be honest with you and tell you that I rely on 2 employees. The rest are not doing much; yes, they do receive all sorts of trainings but without use. All municipal employees have a very lax work ethic. They want to come late, they don't want to work."

This analysis highlights how staffing capacity can be uneven. Within municipalities, the lack of clear delineation of tasks and responsibilities, conflicting and unclear chains of command, in addition to weak staff capacity, especially among the lower tier of municipal staff, is especially debilitating and hinders the provision of quality services to the public.

Interviewed mayors saw the institutional development component of the CITIES activity as being particularly beneficial. Some respondents lauded the technical nature of the training delivered by CITIES. One municipal council member said that the training provided to LDUs is very effective. In Zarqa, the mayor requested CITIES to provide specialized training to municipal units, and CITIES was responsive. At the same time, three respondents commented on what they saw is an overemphasis of CITIES on capacity building. One mayor commented that the training component is more pronounced than the services support component. "What we need is that both components receive the same attention. Training should not trump other needs," he said. In addition, a local council member from the north expressed his reservations about the activity and what he saw is its unnecessary focus on social cohesion. "We have tangible needs and they say they are restricted in what they can do. They have a work plan and need to stick to it."

The theme of staffing, expectations, and ability to address service requests were all common themes within the context of assessing the overall effectiveness of CITIES for achieving its goals of enhancing democratic governance, improving trust and confidence in elected officials, and strengthening stability. The summary provided here highlights the challenges of municipal services and control at baseline. As discussed, the effectiveness of service provision can vary across treatment and comparison groups and within municipalities. Additional data collection rounds will allow the evaluation team to assess what areas or activities may address the staff, perception, and citizen service constraints.

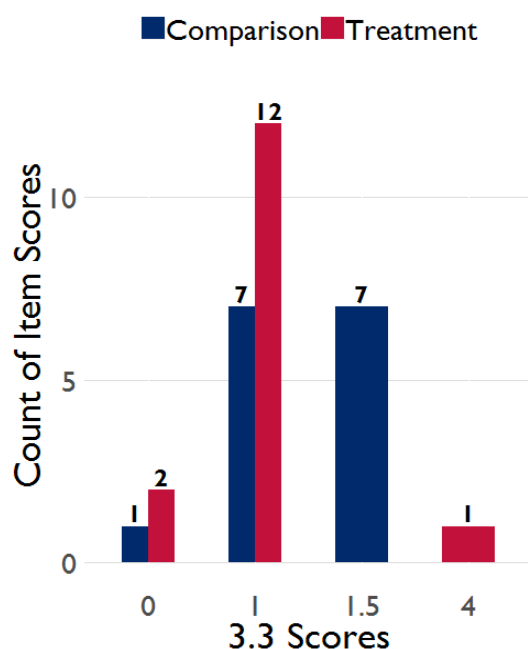
Community Engagement within Municipalities

Municipal respondents reported that community engagement is mostly facilitated through the close personal relationships local and municipal council members have with community members. MICA data

show that only two municipalities in each assignment group are not conducting citizen, community and business engagement to identify needs. Qualitative data collection revealed that council members rely on inspection rounds they undertake in their communities, Facebook pages, and direct contact with citizens who approach them when the need arises. For those municipalities with different districts, citizens engage directly with municipal staff at the main municipality building or with local council members at municipal buildings within the districts that make up the municipality.

Through the 2015 Municipalities Law, the government mandates that municipalities, through their LDUs, engage community/tribal leaders, youth, and civil society to identify priority needs, and engage these groups in the development of the municipality needs guide. The needs guides/manuals that were prepared in almost all municipalities the evaluation team visited cover both municipal services as well as capital projects to be processed at the governorate level. The needs guide effectively covers services and projects that concern municipalities, as well as those that could be supported if endorsed by the governorate council. An executive director said that the municipality includes both sets of needs in the guide, but that the municipal services are only included to inform the executive councils of those needs. The filtering process, according to him, is undertaken by the executive council. Another respondent said the filtering process is undertaken by the LDU at the governorate level.

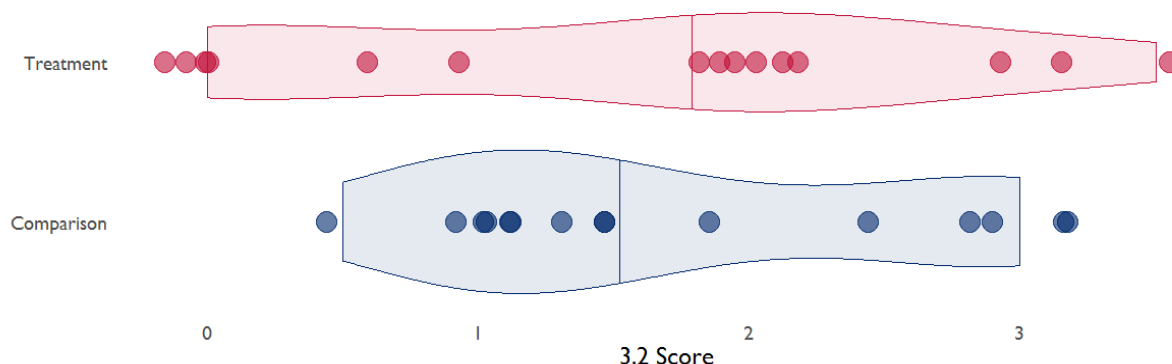
FIGURE 7: MICA ITEM 3.3 SCORES BY ASSIGNMENT



In the MICA tool, item 3.3 records whether or not a municipality conducts citizen engagement to understand local needs. A score of 1 for item 3.3 suggests occasional engagement, while a score of 4 is defined by a municipality that “concretely addresses at least one citizen and/or community need.” As the figure to the left shows, there is at least a minimum level of engagement within the evaluation’s sample across both assignment groups. According to qualitative interviews, the needs guide process has become a central mechanism through which to engage communities. Without other forms of systematic engagement, most interviewed municipalities held three meetings to determine needs. However, according to a CBO head in Tafiheh, the process to prepare the needs guide did not engender significant participation. In one of the meetings, 60 community members attended of which only 10 were women.

Further within the MICA, community engagement through the LDU to produce local development plans, is a key aspect of receiving a higher score (consultation with the community to develop and plan will result in a score of 3). As shown in the figure below, three municipalities in the comparison group scored a three, while two treatment municipalities scored a three and one treatment municipality scored a 3.5. No municipality received a score of four or five.

FIGURE 8: MICA ITEM 3.2 LDU PLANNING SCORES BY ASSIGNMENT



While participation is important for ensuring that citizens provide inputs into local governance, without tangible results, municipal respondents worry that community engagement will backfire. The dearth of resources and the lack of investment projects were mentioned repeatedly as a cause of concern for continuous engagement. Local council members are worried about having to collect needs without the mandate to deliver services or implement projects. The 355 new local councils from 82 municipalities have been voted in. With no resources to deliver services, one mayor stressed that regular community engagement by the councils might backfire if funding does not materialize to cover the identified needs.

In addition to tangible results, accountability is necessary so that public officials explain and justify their decisions to the people. To hold officials accountable, however, citizens must be aware of what officials are doing. Decisions taken during municipal or local council meetings are recorded following meetings, but these records, according to three respondents, are not available to the public. One CBO representative commented that the problem is the lack of transparency associated with municipal and local councils' work. "They could have done something, but the community is unaware of their achievements, if any, because their action plans are not made public." The learning agenda summary in Annex 9 provides additional detail from the general population survey data.

The two interviewed CBOs said that municipalities do not initiate engagement with civil society. CBOs proactively invite municipal staff to participate in their events or occasionally engage municipal council members in discussions about community needs. Both organizations do not lobby municipal representatives to be responsive or monitor their performance. Moreover, CBO representatives do not see their organizations functioning as a watchdog over local government. Their role, as they see it is to address needs in the community through training, charity or other basic support initiatives. One of the interviewed non-governmental organization's (NGO) representatives said civil society outreach efforts to engage municipal staff is ineffective because the mayor does not value the role of civil society.

Learning Agenda Insights

- Knowledge of local government functions is not a strong predictor of governance indicators responsiveness, effectiveness, and confidence.
- Meeting attendance may help improve perceptions of local and municipality councils more so than visits with local officials, social media engagement, or other forms of outreach

It should be noted that the municipal organizational structure has not been reconciled with the 2015 Municipalities Law. While the law mandates that municipalities should engage civil society, there are still no municipal units or departments to take on this role in a sustainable fashion.

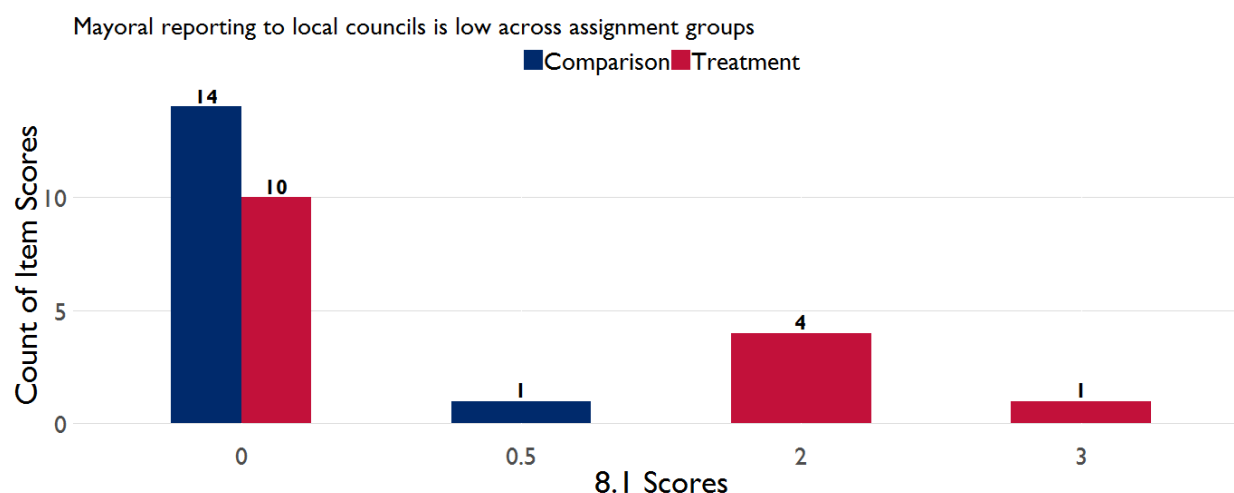
Role of Local Councils

Just a year ago, in August 2017, local councils were elected for the first time. The law mandates local councils to engage citizens, determine their priorities and channel them upwards so the needs can be processed, filtered, and funded. The new structure is meant to enable municipalities to be more responsive to citizens' priorities and better placed to address them.

Council members interviewed stated their role is unclear, and the law does not clearly chart their role or responsibilities. Respondents complained of an unclear mandate and a lack of local budget allocation, which limits their autonomy and ability to act. Local council members said they have little authority to make decisions or effect change. One council member stated, "I do not have the power to make one machine move." Another council member used a similar analogy by saying: "We don't even have the authority to open a fridge. I thought our role would be different...We are a liability on the municipal council, the mayor, and citizens." According to qualitative respondents, this lack of clarity in the delineation of responsibilities creates overlap and redundancy with municipal units and council. For example, one executive director said, "The law did not clearly delineate the mandate and power of the district directors and local council members. The position of District Director is not stipulated in the law so there is clear overlap with Local Councils." In another municipality, the Executive Director complained of continuous friction between Local Council members and District Directors.

In all but one of the comparison municipalities there was no mayoral reporting to local councils, and 10 of the treatment municipalities faced a similar situation. The local council members' confusion voiced in the above quotes is widespread and seen in the figure below, which shows the distribution of MICA item 8.I scores. This MICA item measures the planning and reporting between the mayor and local council. A score of zero implies no regular reporting, while a score of two, which four CITIES municipalities received, suggests that the mayor reports to the council "half yearly," i.e. every six months. A score of three is similar to a score of two but given only if the reporting allows for comparison to earlier reports. In the case of one comparison municipality, Ain Albasha, the executive director reported that he twice prepared reports about municipality achievements and needs, which he then submitted to the mayor, who in turn shared the reports with the municipal council, but this was not a regular occurrence.

FIGURE 9: MICA ITEM 8.1 SCORE DISTRIBUTION BY ASSIGNMENT



Commenting on community members' expectations from local councils, one council member said: "Community members are two kinds: Either they have very high expectations for us or they are smart enough to realize that we have no power and therefore ignore us." Local council members complained of having no designated offices, staff or a separate budget and of not having the power to follow up on the decisions they make in the councils.

Mayors and executive directors also complained that despite municipality's increased financial responsibilities related to the creation of new local councils, governmental allocations have not witnessed a commensurate increase. Most municipal staff perceive local councils as a "liability" imposed on municipalities without a clear assignment of role, staff or budgetary allocations. An executive director said: "Local councils are supposed to drive local development but they do not. Instead their decisions are about what to pay members for meetings, or moving one employee from one municipal unit to another. They are not doing any strategic work." Another municipal council member said: "The local councils are draining the little resources we have...We had to rent out venues and furniture without having seen any increase in our budgets."

Among the other factors compounding the municipal sector's poor financial situation is that municipalities have to now bear the expenses associated with the newly created local councils. Sometimes these are covered through budget items essential for service provision. "Our budget was not increased after local councils were elected. We took money from other budget items like projects and procurement to be able to pay the members of the local councils," a mayor commented.

Role of Governorate Councils

Governorate council members agreed that decentralization is a welcome and timely reform initiative that will ensure the "equitable distribution of the benefits of development." At the same time, similar to local councils, governorate council members interviewed stressed the limited ability of the councils to effect change, their meager resources and the centralization of all efforts related to overseeing the implementation of decentralization. For example, the head of a governorate council stated that the process the executive council undertakes to filter through municipal needs and determine capital projects is still within the control of central ministries. According to him, this undercuts the ability of governorate council members to perform their job. He emphasized that power needs to be devolved from the ministries to

the directorates. The evaluation team's respondents also reported that communication with ministries is weak. Many said that devolving power to the directorates would ensure a closer working relation with the governorate councils.

Governorate council members noted that the development of strategic plans by governorate LDUs should have preceded the development of the needs guides. This would have better facilitated the reconciling of priorities and needs and helped prioritize projects that align with the plans. Council members also noted that regulations clarifying the grounds on which to reject or endorse projects is lacking. "There are demands made by citizens we do not know how to channel. We do not know if these demands should be covered through the needs guide or the government. For example, what do we do about the need for a police station?" one governorate council member said. As alluded to earlier, the needs guide lumps together those services and projects to be provided by municipalities as well as through the governorate council. As one executive director said, all projects are included in the needs guide whether the projects will be covered by the municipality or at the governorate level." Governorate council members expressed their concern that if community demands captured by the needs guides are not met, the public will blame the councils.

Another challenge facing governorate councils is that the 2018 governorate budgets were approved without the input of governorate councils. The few governorate council staff that exist are often employees of the Ministry of Interior (Mol). Their time is divided between their original positions and their newly assigned responsibilities serving the governorate councils. All qualitative respondents emphasized the need for additional regulations to clarify legal provisions that have not served to carve out clear responsibilities and roles for the various councils.

Municipal sector respondents described unsystematic interactions with governorate council members. Respondents reported different forms of engagement by governorate council members in the development of the needs guides, with only a few municipalities reporting close involvement. Municipal respondents also reported limited or no linkages between municipal and governorate level LDUs. Only one executive director interviewed has attended an executive council meeting so far.

One head of a governorate council stressed that decentralization requires that government officials believe in the need to meaningfully transition power to local government. While it would not be surprising for governorate council members to run into bureaucratic resistance to decentralization, the head of a governorate council noted instead the continued application of laws that are contradictory to decentralization's legal framework. According to the same respondent, directorate heads insist these laws are still applicable and binding.

According to three government officials, the implementation of decentralization has exposed gaps in related legislation that need to be addressed. Laws need to be revised and new regulations have to be developed. One reason for that is the opaque mandates of the various layers of local government structures that affect their ability to perform. One respondent stated that the revision of laws is central to this process and must be followed by a focused effort to activate municipalities' developmental role. According to this respondent, municipalities are still focused on the delivery of conventional services when they should be pursuing development opportunities to uplift their municipalities. The engagement of the private sector is key to this new role as well as the need to build LDUs' capacity to lead on development opportunities.

One government official also emphasized the need for a strategy for spatial planning as a form of national urban planning to enable governorate council members to make governorate-level decisions and prioritize projects in each governorate. The same government official said that municipalities must prioritize projects that propel local development and provide employment opportunities.

Perceptions of Decentralization

Respondents reported that the new system has ensured the election of representatives to new tiers of local government. It has also assigned new responsibilities to what most respondents saw as inexperienced actors. By the admission of municipal respondents, existing challenges include the new council members' weak capacity, lack of experience in public service and the heightened expectations of community members who expect decentralization to yield short-term wins. One CBO member noted that many citizens believe government appropriations to governorates have increased as a result of the recent decentralization process, and therefore expect significant changes.

According to most respondents, the new decentralized system is fraught with challenges. While they agree that elected representatives standing for the new administrative layers are closer to the people and can better understand and engage community members, they insist that decentralization has not yet produced tangible changes and will need time to take root. "Decisions are still made by the ministries," one mayor commented. Respondents also emphasized the ambiguity of the law and the overlap in mandates. "We don't know if the governorate council members belong to the decentralized system or the centralized system," another mayor commented.

With resource allocations for both governorates and municipalities unchanged, five municipal respondents negatively assessed recent changes saying the situation is regressing with decentralization serving to worsen the ability of local governance structure to address communal challenges and needs.

Another challenge relates to the lack of engagement between the layers. According to all respondents, municipal LDUs are generally weak and collaborate little with their counterparts at the governorate level. Governorate council members also reported the lack of engagement with executive councils.

According to three government officials, CITIES is attracting some criticism related to the project's capacity building efforts within the context of decentralization. Criticism is centered on the lack of practical knowledge CITIES trainers have, the time allotted for training, and what is perceived as weak follow up to the training. During an interview at MoMA, the government representative stated that trainers lack the practical knowledge to respond to questions about the implementation of decentralization. The respondent added that one-day workshops are seen as inadequate to build capacity. "These are complicated issues. One cannot cover the areas they are trying to cover in one day. It is just not enough. How can people digest all this content in one day?" the ministry official said. The same official expressed concern about what he regarded as CITIES' focus on capacity building and training:

"So many organizations are training. CITIES is training the same people that MoI, the Electoral Commission and USAID Community Engagement Program (CEP) have trained before. How can this be effective? How does one measure the impact of all this?"

The government official went on to comment on the lack of follow-up to the training to assess if trainees are applying the acquired learning. Training, according to respondents, should be focused on the practical application of the legal infrastructure. Respondents noted that training must be applicable in the context

from which the various council members hail. “The training we are being offered tends to be general when it needs to be focused and highly technical,” another central ministry official said. The head of one of the governorate councils added that while training is necessary, it is also useful to share experiences with those who have gone through decentralization so that council members can understand how the system is supposed to function.

Government officials at MoMA expressed other concerns about the CITIES project. One official stated that CITIES did not provide a work plan even though the ministry had requested it several times. Another official said the Minister is requesting a review of the training content CITIES is using to train local, municipal, and governorate council members. He added that the implementation of the project has generally been slow.

SUMMARY OF EFFECTIVENESS FINDINGS

The quantitative and qualitative data collection conducted through implementation of the MICA tool and in-depth interviews with municipality stakeholders provide a snapshot of a local context in which the following observations can be made:

- Municipal capacity to strengthen local service delivery is limited overall due to limited autonomy, staffing, and resource constraints in both assignment groups;
- Community engagement mostly relies on direct contact and personal relations. It is slowly being operationalized across both treatment and comparison groups. There is lack of systematic feedback or use of citizens input for planning, and there are fears of raising expectations that municipalities cannot meet;
- Local councils have no resources and an unclear role across both assignment groups. There is almost no reporting happening between mayors and the municipal councils in the comparison group;
- The governorate council is still determining its role and relationship with the central ministries and municipalities, with respondents noting the limited devolution of power, the need for legislative reforms and policy guidance to implement the decentralization law;
- There remains confusion around decentralization, gaps in the legal framework, weak engagement and linkages between governmental layers, and continued centralized decision-making. Respondents report mixed perceptions of the effectiveness of CITIES’ support for capacity building around decentralization.

The above findings highlight important dynamics that will likely have an effect on CITIES’ effectiveness. How effective CITIES is at addressing these factors will become more apparent during subsequent rounds of data collection. As observed by several respondents, follow-up on training and capacity building that strengthens municipal functional and technical abilities may be critical for effectively promoting service provision, needs assessment, and citizen engagement, as well as the various efforts to promote decentralization and local capacity.

EVALUATION QUESTION 2: SUSTAINABILITY

More time will be needed to determine the sustainability of specific CITIES interventions, but a number of factors and risks to sustainability emerged. These include:

- Endorsement of CITIES' system level initiatives by the national government
- Institutionalization of decentralization
- Municipal capacity constraints and control
- Harmonized donor capacity building support strategies
- Civil society capacity to fulfil its role in municipal accountability

Going forward, the sustainability of CITIES interventions will depend on whether system-level changes instituted by the activity are endorsed by the national government and whether municipalities and the government are capable and willing to exercise ownership of activity-facilitated reforms. For example, CITIES supported governorate development plans and the methodology by which they were developed will need to be endorsed by governorates to ensure that current and subsequent councils refer back to the plans and engage in their implementation. The same applies to municipal community engagement mechanisms or bodies, new organograms for MLDUs (one each for Category A, B, and C municipalities) and job descriptions of MLDUs' staff. Without official endorsement or recognition, financial resources will not be allocated to sustain these mechanisms.

The GoJ's slow implementation of reforms or devolution of power needed for municipalities to assume new responsibilities and the slow or inadequate provision of financial resources to municipalities or governorates may affect the sustainability of intervention benefits. Capacity constraints, lack of incentives and weak motivation may also deter municipal staff participation and the adoption of needed corrective actions. Sufficient resources are needed to implement municipal local development plans and governorate plans. Resources are also required to address the needs identified through municipal community engagement and to incentivize staff to perform and adopt changes. Also, the lack of municipal ability to recruit qualified staff may affect municipalities' overall ability to provide services. While municipalities have expressed the need to hire more qualified staff, they may not be able to secure the talents they need. MoMA has final say as to recruitment of new employees and capacity building in the form of training alone will not be adequate to close qualification gaps in current staff.

The harmonization of donor support especially when it comes to the provision of capacity building assistance will be critical. Duplicative and unharmonized donor programming could lead to municipal staff adopting different procedures and processes to undertake the same municipal tasks. At the same time, the nature of capacity building support whether it emphasizes the transfer of technical knowledge or only awareness raising may affect how it is received in the various communities. In this context, a capable and empowered civil society able to monitor continued implementation of reforms and overall municipal performance as a guarantor of intervention continuity and value can be instrumental in ensuring sustainability.

Most importantly, public perception of what constitutes municipal services may also affect project impact as citizens' inflated perception of municipal responsibility sustains their heightened expectations of what the municipality can and cannot do. As findings confirmed, citizens conflate the role of their municipality with that of the government, expecting it to provide services beyond its capacity or mandate. Several municipal staff stated that citizens expected municipalities to service main roads or address water issues they are not responsible for.

EVALUATION QUESTION 3: SYNERGY

Baseline data suggests that there has been some initial coordination between CITIES and other donor activities, but there are salient gaps in overall donor coordination and synergy to overcome. CITIES continues to pursue collaboration with donors and implementing partners with shared or complementary objectives to support decentralization and local governance. Through this baseline study, over 28 implementing agencies operating through 48 projects implementing 208 separate interventions at the central and local levels were identified. Speaking to various donors about the nature of their support and how it fits with government's efforts, donors agreed on the existence of an overall national vision for change but emphasized that a clear roadmap for how to bring this change about is still lacking.

The MoPIC is responsible for the overall coordination efforts between donors. According to government officials, MoPIC's coordination efforts are focused on the national and policy levels without combing through the details of specific projects. On the other hand, MoMA provides donors working in the decentralization space with guidance, coordination, and facilitation support. According to donors, what partly hampers effective coordination is different MoMA staff overseeing donor projects with little horizontal coordination between them. Compounding this situation according to a government official is that projects can reach municipalities without first passing through MoMA. According to the same official, municipalities are not under any obligation to publicize donor programming in their communities and government officials cannot force them to divulge information about such programming.

Interviewed government officials agreed that coordination efforts are unsystematic. This is due, according to one official at MoMA, to the lack of an overall framework that connects disparate components of programming. "When it comes to decentralization, we only have laws. There isn't a framework for our work. Efforts are not systematic. We are trying to coordinate and connect disjointed interventions." According to the official, the drive for decentralization should have started with the development of an overall policy delineating the government's vision for the initiative, followed by the introduction of laws and subordinate legislation then strategies to carry out the envisaged change. This long-term vision "would have prevented the chaos we see today...We are trying to rectify the situation and learn from our mistakes," the official said.

The EU is currently working on a strategy for decentralization. The strategy is part of the Decentralization and Local Development Support Programme (DLDSP), a joint project of Local Development Directorates (LDD) of MoI and MoMA. The DLDSP is funded by the EU and implemented by the UNDP. The program provides support to the GoJ for the design and implementation of decentralization-related reforms and the building of sub-national governance and administration capacity. Support provided is helping the GoJ guide the decentralization and local development process and develop the relevant institutional and organizational settings for its implementation. The DLDSP operates under the guidance of the Inter-Ministerial Committee (IMC) for Decentralization and its Executive Committee and is supporting the IMC to make the required modifications at both the governorate and municipality levels to ensure alignment with the newly modified mandates introduced by the legal framework. A key component of this work is to support the GoJ to develop its vision and policy to reform local administration and to formulate a National Strategy and Programme for Decentralization.¹⁶

¹⁶ Mapping of Development Partners Interventions in Supporting Local Governance: Local Development in Jordan, 2018, p.6.

The government has not yet approved the development of the strategy and remains in a diagnostic phase as to how reforms are being received in communities. The government is focusing on the governorates of Balqa, Aqaba and Irbid to monitor the implementation of the decentralization-related reforms.

Donors echoed opinions voiced by government officials as to the lack of an overall vision for decentralization. According to respondents, when it comes to approving the development of a strategy, the government is procrastinating with ministers holding different interpretations of what the process of decentralization entails. One respondent stated: “the biggest problem here is that there isn’t one voice or one vision. The respondent added: “We are supporting something, but we don’t know what it is.”

In addition to the IMC, which is the main part of the Jordanian government’s decentralization support structure, the government has established various committees in support of decentralization, including the ones for legislation, resource management, and finances. The committees are all working under the umbrella of the IMC. However, as donors are supporting different parts of government and in the absence of a clear government vision, how donor agendas, understanding and priorities for decentralization are affecting its implementation remains unclear.

Beyond decentralization, municipal staff still emphasized the weak coordination efforts between donors. An executive director in a municipality in the north stated that when it comes to donors, the largest problem municipalities face is the lack of coordination between them. He added: “they come in and insist on focusing on the same area even if the need is already being met by another donor.” Even within the same donor, according to a government official, CEP worked with MoI to develop a guide to lead a community-focused process to identify needs. MoMA officials expressed a preference that CEP work with them first. As a result, CEP worked with MoMA to develop another guide for a similar process at the municipal level. The result is two guides, one geared towards the governorate level and the other focused on the municipal level.

To address challenges in local capacity, various organizations and projects are or will soon be implementing direct and indirect capacity building interventions to support local, municipal, executive and governorate council members including at least 13 different projects spanning across a multitude of donors from the USAID, WB, AFD, EU, AECID, GIZ, and others.¹⁷ According to Jordanian officials, an institutional framework for capacity building has been developed by MoPIC, and the ministry is trying to harmonize efforts to support capacity building.

The government is also in the process of setting up a local governance training institute to build the capacity of council members and municipal staff. A special committee was set up to oversee the institute’s establishment and curricula and harmonize donors’ related efforts. It includes representatives from MoMA, the municipal sector and MoPIC. The progress of the project, as a MoMA official admitted, is slow.

17 CITIES, Local Enterprise Support Project (LENS), International Republican Institute (IRI), National Democratic Institute (NDI), Danish Aid, Federation of Canadian Municipalities (FCM)/ Jordan Municipal Support Project (JMSP), Italian Agency for Development Cooperation (ITDC)/ Support of Municipalities Affected by the Syrian Refugees Influx in Jordan (SMASRIJ), Japan International Cooperation Agency (JICA)/Waste Management Project (WMP), Spanish Agency for International Development Cooperation(AECID)/Decentralization Support Project (DCP), Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ)/Supporting Participatory Resource Management to Stabilize the Situation in Host Communities (PRM), Organization for Economic Cooperation and Development (OECD)/Partnering for Good Governance and Open Government (PGGOG), International Cooperation Agency of the Association of Netherlands Municipalities (VNG)/(Local Governance and Resilience Program (LGRRP) and Agence Francaise de Developpement (AFD)/Regional and Local Development Program II (RLDP II), among others.

Only a few meetings to discuss plans have taken place. According to WB respondents, the WB's new project supporting municipalities also contains a capacity development plan. The plan has not been harmonized with the government's plans to establish the institute. "We cannot freeze what we want to do until the government's own plan is implemented," the WB respondent said.

When it comes to CITIES led coordination, all interviewed donors agreed that CITIES is trying to coordinate efforts. However, the WB believes that the government should be taking the lead on this front. According to WB respondents, coordination is difficult when a "simple matrix on what everyone is doing" is lacking. Commenting on this effort, one donor representative said: "CITIES is trying to coordinate. It is not enough. They held their last meeting last April I think. We exchanged ideas but that's it. We go there to learn what others are doing but not really to coordinate."

MoPIC has in fact requested the UNDP to map all the support being provided at the municipal and governorate levels to assess donors' interventions relevant to needs, identify gaps and avoid duplication. This resulted in a matrix of interventions by development partners in support of the implementation of the new legal framework and more precisely the sub-national system at the governorate and municipal levels. The final matrix identified 28 implementing agencies operating through 48 projects that translated into 208 separate interventions at both the central and local levels and in the following areas¹⁸:

- Policy, Legislation and Regulation
- Functional Assignment
- Institutional and Organizational Structures
- Inter-Governmental (Institutional)
- Systems, Procedures and Manuals
- Capacity Development
- Local Public Finance

Interviews with the WB revealed that CITIES has been coordinating with the new MSSRP project. The WB will be using CITIES' manuals and training material and the activity's consultants to train and roll out in non-CITIES municipalities. According to WB respondents, the two projects diverge on the issue of ensuring accountability for the needs guide process. The WB sees the need for closer monitoring of community engagement to ensure participation is meaningful while CITIES wants the process to unfold more organically. WB representatives said that CITIES manuals would be reviewed. If they are not sufficiently rigorous they will be supplemented in non-CITIES municipalities. CITIES is also coordinating with the WB on the Technical Assistance of Municipal Financial Capacity project to ensure there is no duplication.

In addition, there is an OECD project focused on participatory budgeting and development planning. Respondents said they have met with CITIES several times and are closely coordinating the two projects. The two respondents agreed that decentralization would improve the quality of services if the system is functional and allows needs to translate upwards to be better reflected in budgeting. Otherwise, and because of citizens' high expectations, the system might backfire. There remains a big issue of trust in the links between the administrative layers. Engagement must be inclusive and meaningful. Respondents said

¹⁸ Ibid.

more effort has to go into understanding how grassroots priorities are filtered as they move up to subsequent local government levels; this, according to respondents, boils down to the quality of regulation.

AECID, AFD and VNG respondents also confirmed they are coordinating with CITIES. AECID is focusing on the three municipalities of Mafrq, Sarhan and Ramtha and is working to improve financial management. AFD's project Regional and Local Development Project (RLDP II) is expected to start end of 2019 and focus on the governorate centers of Salt, Ajloun, Madaba, Irbid and Jerash. In collaboration with AED, CITIES helped MoMA to achieve triggers related to the first municipal sector policy loan (SPL) and will continue to do so for the next SPL. VNG and in addition to its capacity building support is partly focusing on Mafrq working in the three municipalities of Balama, Al-Sarhan and Umm Al-Jimal to develop green public spaces.

INTERNAL VALIDITY OF MICA SCORES

Any impact evaluation must address the issue of internal and external validity. External validity generally refers to how applicable the findings of a given study are beyond that study's context. In the case of CITIES, there may be questions about how applicable these findings are to another Middle East and North Africa (MENA) region country, or even to non-study municipalities. A focus on the mechanisms for impact and careful noting of contextual factors can help address external validity concerns for decision makers. Internal validity broadly refers to how well research is implemented. Threats to internal validity can arise if, for example, data collection occurs in a non-uniform way or there is contamination between assignment groups. The threat of data contamination or inability to maintain comparable respondent groups is low in the context of the MICA data—it is unlikely that a mayor from a treatment municipality would somehow respond to a comparison municipality MICA question. The MICA tool itself was applied equally across each comparison municipality by the ET (and in treatment municipalities by CITIES). Although there were subjective differences in MICA scoring between ET enumerators, enumerators asked every question and applied the tool the same way for each municipality. With the exception of one municipality, enumerators conducted MICA interviews together, but recorded their own notes, and scored independently based on their notes.

There are two main threats to the internal validity of the MICA scores:

- Enumerator bias
- Non-comparable MICA application across the CITIES and evaluation teams

The agreement statistics suggest that there is a strong potential for enumerator bias, or the bias that arises from how enumerators ask and record questions, for MICA services items. However, the evaluation team's process for data review mitigated the threat of this bias. As described above in the evaluation design and data collection sections, the ET first identified all errors and reviewed item-by-item differences. The evaluation team then undertook a MICA synthesis process to align scores across raters to develop a final dataset of comparison municipality MICA scores. In some cases, MICA guidance was ambiguous or inconsistent leading to an increased role for subjective judgment in determining a score. As part of the data review process, the enumerators went over their notes for each item where there was disagreement, discussed what they observed, and determined whether to maintain or change the final score.

The CITIES team applied a similar process of scoring and reviewed all final scores with the technical component lead. Despite similar processes, there remains a threat to internal MICA data validity between

the CITIES team and the ET stemming from any differences in the application of the MICA from the ET, or if their review process somehow biases scores. By training, meeting, and consistently communicating with the CITIES team, the evaluation team sought to align its application of the MICA tool with CITIES. Both teams asked each question in the tool, questioned the relevant stakeholders based on CITIES’s guidance in applying the tool, and took diligent notes. Despite these similarities, there are likely differences in MICA application because the CITIES scoring team was assembled based on its expertise in the relevant MICA criteria. This could, in theory, result in an ex-ante difference in their MICA-related measurement skills that training may not be able to resolve.

The MICA data do suggest slightly different usage of the tool. As shown in the figure below, item-level analysis across assignment groups generally shows that the ET used interim scoring more often; for example, giving the score of 0.5 58 times compared to CITIES 19. The evaluation team explored the use of half-scores with the CITIES team to understand when to best utilize these since the MICA tool has clear integer categories. Additional discussion with CITIES staff may be warranted between this round of data collection and the next implementation of the MICA in the comparison municipalities to ensure that the evaluation team is properly aligned with the MICA methodology.

Figure 10: Distribution of MICA Services Item Scores by Assignment Group



The next section addresses scoring equivalence across assignment groups, but understanding how the MICA tool and subsequent scores were functionally applied is important for understanding threats to the internal validity of the data itself. Beyond using half-scores more frequently, there is no clear evidence that the tool itself was applied in a different capacity across treatment and comparison groups. On an item level, there is some clustering around certain values, such as item 7.2, but the notes of the MICA tool suggest that this clustering, as well as the evaluation team’s use of half-scores is consistent with guidance from the CITIES team.

The threats of enumerator bias and incomparable MICA application are real, but resolvable threats to the internal validity of the MICA data. The evaluation team addressed these two threats through its own internal reliability processes and through its training and outreach with the CITIES team.

INTERNAL SCORER AGREEMENT ERROR

There is also a concern about measurement consistency with any data collection effort. Two reasonable people can interpret the same question in slightly different ways, which can affect the reliability of the overall measurement endeavor. In the case of the CITIES IE, the use of the MICA as a primary data source for this IE presents two main challenges: between team reliability and within team reliability.

The evaluation team's MICA scoring in control municipalities needed to be aligned with that of CITIES' scoring in its treatment municipalities to ensure between-team comparability. The evaluation team and CITIES needed to have a similar understanding of each MICA item and how to apply MICA scores appropriately. To this end, the evaluation team received two training sessions with members of the CITIES team. The ET also shadowed CITIES staff as they applied the MICA in six Mafraq governorate municipalities. The goal of training with and shadowing CITIES staff was to minimize the variability between each group's scores and overcome the challenge of each team implementing the MICA tool differently.

During the training, members of the CITIES team were told that the tool could produce socially desirable answers or impressionistic responses that may not hold up to objective scrutiny. The CITIES team emphasized the importance of requesting municipal staff to produce evidence in the form of related documentation when responding to questions. The ET followed the guidance but was still able to identify ratings within the tool that covered subjective criteria or rubrics that did not exhaustively define applicable objective criteria and therefore allowed evaluators to exercise their discretion and assign a higher or lower score depending on their opinion or particular bias.

The second challenge MICA scoring presented as part of baseline data collection was within-team, or inter-rater, reliability. Even though the ET received training on how to employ the MICA tool, there could still be variability within the ET on how to appropriately score the same municipality on the same measures. Beyond the training and guidance that the evaluation team received from CITIES, the MICA tool itself has guidance on how to apply scores for each item. For example, on MICA measurement 2.1, Technical Trainings in Service Delivery, the MICA tool provides the guidance to apply a score of 1 if the municipality occasionally organizes training on technical topics (e.g. maintenance infrastructure, maintenance of vehicles/equipment, SWM), at request [sic]. This seems straight forward, but there are often cases where conditions on the ground are more ambiguous, such as whether staff were actually able to provide services and their qualifications to do so. If there is a lot of variation within the ET on what constitutes one score over another, assessment of the comparison municipalities could be biased as would comparison to CITIES' treatment municipalities.

The table below provides a summary of agreement between evaluation team enumerators across each municipality, as well as a set of municipal-level agreement statistics (the intra-cluster correlation coefficient statistic) most appropriate for ordinal level data. This analysis was part of the overall reliability and scoring process that took place after baseline MICA data collection was completed.

Table 6: Total MICA Score Reliability Statistics

Municipality	Percent Agreement	ICC
Al Seru	73.91	0.84
Alsharah	73.91	0.83
Ain Albasha	60.87	0.88
Alauion	60.87	0.69
Talal Aljadedah	60.87	0.75
Alyarmook Aljadedah	56.52	0.87
Bab Amman	56.52	0.84
Manshiat Bani Hasan	56.52	0.73
Shafa	52.17	0.78
Mo'ath bin Jabal	47.83	0.70
Tabaqat Fahl	47.62	0.64
Iel Jadedda	43.48	0.75
Jizah	43.48	0.30
Mazar Jadedda	43.48	0.80

Agreement ranged from 43 to 74 percent. A percent agreement of 100 would result if all of the enumerators had identical MICA scores for a given municipality, whereas, the 56 percent agreement for Shafa, for example, suggests that there was disagreement on almost half of the MICA items. It is important to note that percent agreement on its own may not be a fair assessment of enumerator alignment. One limit of this approach is that it simply measures whether enumerators assigned scores identically, but it does not account for the fact that scores could be similar, but not the same, among raters for a given municipality. The intra-cluster correlation coefficient (ICC) provides a measure of how similar scores are within a given municipality, or “cluster.” A higher ICC suggests a stronger relationship between scores for a given set of raters within a municipality; for example, if the ICC were 1.0 for a given municipality, all of the raters would have identical scores. Mazar Jadedda provides a good example of how the percent agreement and ICC differ. The percent agreement for Mazar Jadedda is 43.5, suggesting that raters’ scores differed for about 56.6 percent of the time, yet the ICC is 0.80, which suggests that although the raters’ scores differed, they were very closely related. The item-level scores for Mazar Jadedda differed by about 0.47 on average, with only three items differing by more than one point (i.e. the magnitude of the overall difference was low); in contrast, in Jizah, which has the same percent agreement but an ICC of 0.30, scores differed by an average of 0.52, including five items with an item-level difference larger than one.

To understand error more directly and descriptively, the evaluation team manually calculated the error rate, i.e. the rate at which one or more evaluation team members scored the same item for the same municipality differently. As shown in figure 13, the error rate varied across municipalities. Annex 3 provides more detail and additional agreement statistics such as for continuous data.

FIGURE 11: TOTAL MICA ERROR BY COMPARISON MUNICIPALITY

A review of a simple percent agreement statistic, examining a range of agreement statistics customized to various data types (categorical, ordinal, continuous), and a manual exploration of total error across raters, municipalities, and items indicates the existence of a low to moderate dissonance across ET MICA scorers. This dissonance varies but was more pronounced in the case of a few municipalities and a few items.

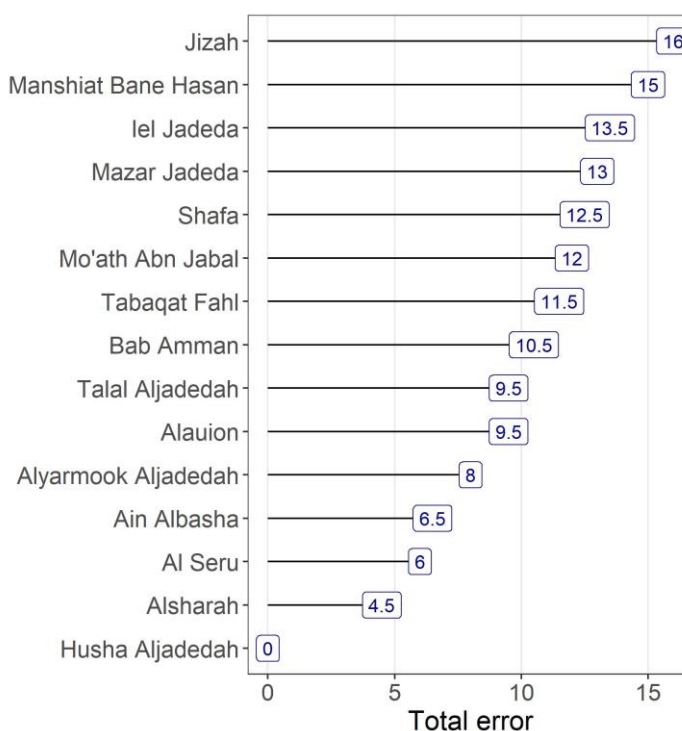
As mentioned above, to address this internal error, the ET conducted an internal review and synthesis process, as well as a review with the CITIES Component I lead, to determine final harmonized scores to address internal errors and mitigate reliability concerns. For any areas where MICA scoring guidance was unclear or ambiguous, the

team agreed upon a unified way to score that particular MICA element across all municipalities. For example, under element 1.2 “Manual Records and Licensing System,” the MICA scoring rubric lays out a progression whereby all permit and licensing forms, records, documents, and procedures are prepared for migration from manual form to a database system prior to any migration. The highest score is awarded when the municipality is ready to migrate all forms, documents, and procedures to a database system. In practice, the ET observed that municipalities digitize in stages, often prepare and digitize certain licensing or types of records first, and then move on to prepare and digitize others. After discussion, the ET team agreed to award any municipality that had partially automated their licenses a “one.” A municipality nearing completion automating all their licenses would receive a score of “three.”

For any scoring issues that raised questions around internal validity between CITIES’ scoring and the ET, identified dissonance was reviewed in conversation with the CITIES Component I lead, and ultimately considered within normal bounds and the comparison municipality scoring to be sufficiently valid as to compare against the CITIES MICA scoring.

ASSESSING BASELINE EQUIVALENCE

If the MICA scoring in comparison municipalities is sufficiently valid as to compare against the CITIES municipal scoring, the natural next consideration is the difference across assignment groups. That is, how



do the CITIES municipalities compare to a set of similar municipalities that do not receive CITIES implementation? The question is critical for maintaining validity of the evaluation design - if substantial differences across municipalities exist, the control group may not be a valid representation of the counterfactual, or additional weighting may need to be applied to improve the comparability.

As outlined in the evaluation design document, the evaluation team conducted an initial selection of comparison municipalities through an ex-ante matching process to create observably similar assignment groups on general municipal characteristics.¹⁹ While the initial municipality matching procedure sought to ex-ante increase the probability of balance between the treatment and comparison groups, it is important to test this assumption once data are collected to confirm the fidelity of the procedure and see what, if any, additional steps may be needed before the next stage of analysis. With scoring of the comparison municipalities concluded, it is now possible to directly assess baseline equivalency across municipal institutional measures.

Note that balance metrics are assessed against three configurations of treatment and comparison municipalities. The first configuration is all 33 CITIES municipalities against the 15 comparison municipalities. Given that this configuration includes Class A municipalities that are entirely a part of the CITIES activity, it is for reference only and not a central part of the evaluation.²⁰ The second configuration is the 23 CITIES municipalities that exclude class A, compared against 15 comparison municipalities. The third configuration consists of 15 treatment municipalities compared against 15 comparison municipalities.²¹

The value of assessing equivalency across multiple configurations is that the evaluation team can better understand how sensitive scoring is to particular metrics or municipality characteristics. It may be the case that the matching of 15-15 was overly conservative based on similar data, but not the exact metrics of interest for this IE. By testing the three aforementioned configurations, the evaluation team can assess whether and by how much scores across assignment groups differ. This is important for understanding how robust the findings are and whether they may carry over to a broader implementation context.

The balance metrics that follow are organized by aggregate and item-level scores across score type (Services and GESI), and across the three configurations mentioned above. Finally, each of these set of metrics is presented across the final set of scores produced by the evaluation team, and the alternate set of scores after harmonization.

AGGREGATE BALANCE

The MICA score is predicated on the concept of an aggregate score of municipal level competency consisting of the sum of 23 services and 17 GESI scores. The CITIES evaluation, specifically Hypotheses I-3 (municipal institutional level outcomes), takes the further step of supposing an overall capacity measure

19 After removing the ten class A municipalities, class B-D municipalities were matched on the criteria of overall and refugee populations, poverty rate, and presence of either CEP or ESSRP. A set of 15 comparison municipalities were selected that, in aggregate, were most closely balanced against the set of 23 treatment municipalities – again in aggregate. From this initial specification, the matching process was re-run to select 15 out of the 23 treatment municipalities that represented the closest aggregate match to the 15 comparison municipalities.

20 In the event that the evaluation finds a CITIES treatment effect among class B-D municipalities, the chief interest in the first configuration will be to observe the extent to which such an effect may replicate among class A municipalities.

21 Ideally, any CITIES treatment effect will persist across each of these two configurations.

consisting of the overall mean MICA score across the entire set of treatment municipalities under study, compared to the overall mean MICA score across the entire set of comparison municipalities.

To calculate final aggregate balances, and support any subsequent sensitivity analysis, two sets of scores are presented below and analyzed for baseline equivalence. The first set of scores, labeled “Final MESP MICA Scores” utilizes comparison municipality scores that were agreed upon by the ET prior to any adjustments made in conversation with the CITIES Component I Lead. A second set of scores, labeled “Alternate MICA Scores” is also presented, which reflect any adjustment in scores made as a result of scoring review conversations with the CITIES Component I Lead as part of the process to strengthen inter-team scoring reliability described in the evaluation design and inter-rater reliability sections above. While both sets of scores are presented, primary analysis informing conclusions are from the “Alternate MICA Scores” unless explicitly noted otherwise.

Overall balance indicates a mixed picture of achieving baseline equivalence depending on what municipalities are included in the analysis and the MICA items assessed. More detail is provided in Annex 4, but a standardized difference with an absolute value of 0.25 or less indicates particularly strong balance, while a value at or above 1.0 should raise serious concerns. The Selected Municipalities and Municipalities Excluding Class A were balanced across services items, which is valuable for the evaluation since much of the focus of the MICA tool is focused on services items and the 15-15 matched municipalities are the target groups for this evaluation. However, as seen in the table below, aggregated balance is less apparent for the GESI items and for the broader CITIES-comparison municipalities for “Final ET MICA Scores,” i.e. the scores the ET collected and that were not updated based on conversations with CITIES.

Table 7: Final ET MICA scores by treatment, selection status

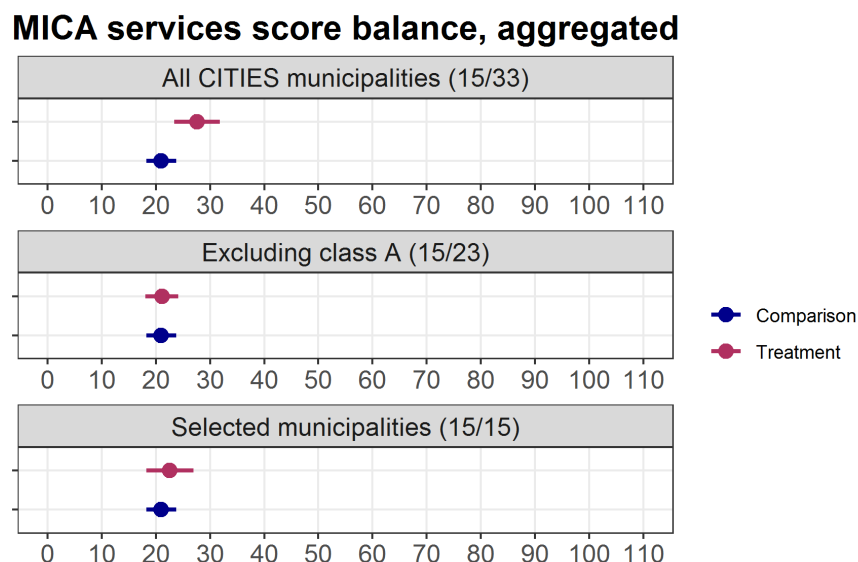
Category (Treatment/ Comparison)		Component	Comparison Mean	Treatment Mean	Difference	Standardized difference
All municipalities (15/33)	CITIES	Services	21.07	27.62	6.55	0.54
Excluding class A (15/23)		Services	21.07	21.11	0.04	0.01
Selected municipalities (15/15)		Services	21.07	22.57	1.5	0.18
All municipalities (15/33)	CITIES	GESI	0.53	1.79	1.25	0.74
Excluding class A (15/23)		GESI	0.53	1.52	0.99	0.7

Category (Treatment/ Comparison)	Component	Comparison Mean	Treatment Mean	Difference	Standardized difference
Selected municipalities (15/15)	GESI	0.53	1.53	1	0.75

For MICA services scores summed across municipality for an overall services capacity score, and then municipal scores averaged by treatment or comparison status, there is equivalence for the specification of 23 CITIES municipalities compared against 15 comparison municipalities (standardized difference .02). For the specification of 15 CITIES municipalities and 15 comparison municipalities, the standardized difference and raw mean difference suggest statistically similar aggregate scores. For all 33 CITIES municipalities compared against the 15 comparison municipalities, the standardized difference of 0.54 is unacceptably large, and indicates substantive differences between groups. This substantive difference is at least partially observable in that Class A municipalities are generally of higher functional capacity and better resourced.

Differences in aggregate services scores are visualized below.

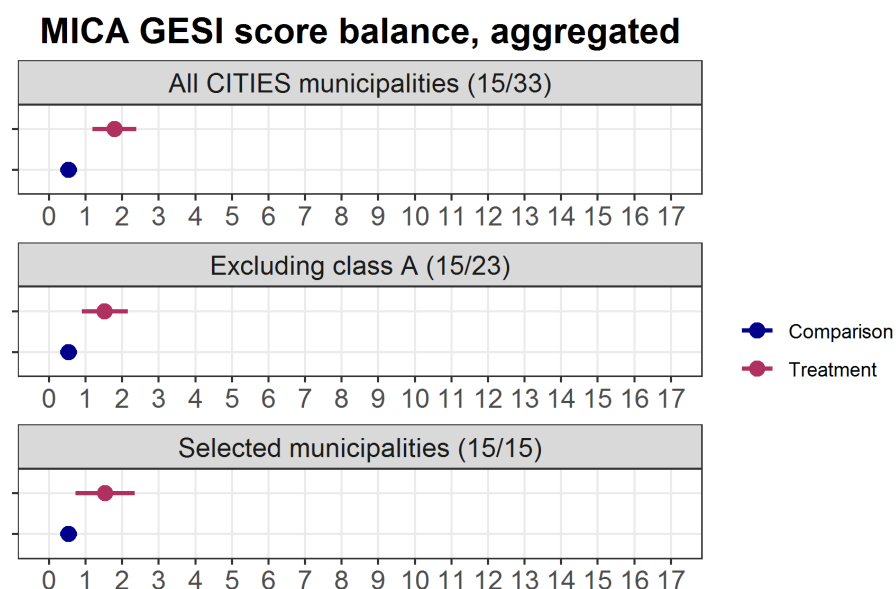
FIGURE 12 OVERALL FINAL ET MICA SCORES BY TREATMENT, SELECTION STATUS



For MICA GESI scores summed across municipality for an overall GESI score, and then municipal GESI scores averaged by treatment or comparison status, balance metrics are unacceptably high. However, at least some of the poor balance properties for GESI scores may be considered to be floor effects – scores are very close to zero and objectively small differences become large relative differences. Related to the floor effect, the binary nature of the GESI scoring system could possibly be influencing scores close to zero. To receive a score of one, a high threshold needed to be met, as compared to the more gradated scoring system for services scores. Even so, the feasibility of using GESI scores is called into question, and

any estimates generated from GESI scores should be scrutinized carefully. Differences in aggregate GESI scores are summarized below.

FIGURE 13 OVERALL FINAL ET GESI SCORES BY TREATMENT, SELECTION STATUS



Here differences across treatment and comparison status are more apparent and could be related to outcomes.

The following table presents the same aggregated balance, but this time using the alternate scores.

Table 8 Alternate MICA scores by treatment, selection status

Category (Treatment/ Comparison)		Component	Comparison Mean	Treatment Mean	Difference	Standardized difference
All municipalities (15/33)	CITIES	Services	22.6	27.62	5.02	0.41
Excluding class A (15/23)	class A	Services	22.6	21.11	-1.49	0.19
Selected municipalities (15/15)		Services	22.6	22.57	-0.03	0
All municipalities (15/33)	CITIES	GESI	0.73	1.79	1.05	0.61

Category (Treatment/ Comparison)	Component	Comparison Mean	Treatment Mean	Difference	Standardized difference
Excluding class A (15/23)	GESI	0.73	1.52	0.79	0.54
Selected municipalities (15/15)	GESI	0.73	1.53	0.8	0.57

Under the alternate scores the balance metrics on the services scores improve, but there is also a substantive change. Under the final internal scores, there was equivalence in the 15/23 specification and the 15/15 specification was marginally acceptable. Under alternate scores the situation is reversed: the 15/23 specification now has a standardized difference of 0.2 while the 15/15 specification has full equivalence across treatment and comparison status. If a choice is to be made between alternate and final scores, the ET recommends proceeding with alternate scores as a basis for midline and endline data collection, but comparison across both MICA datasets should occur as part of sensitivity analysis.

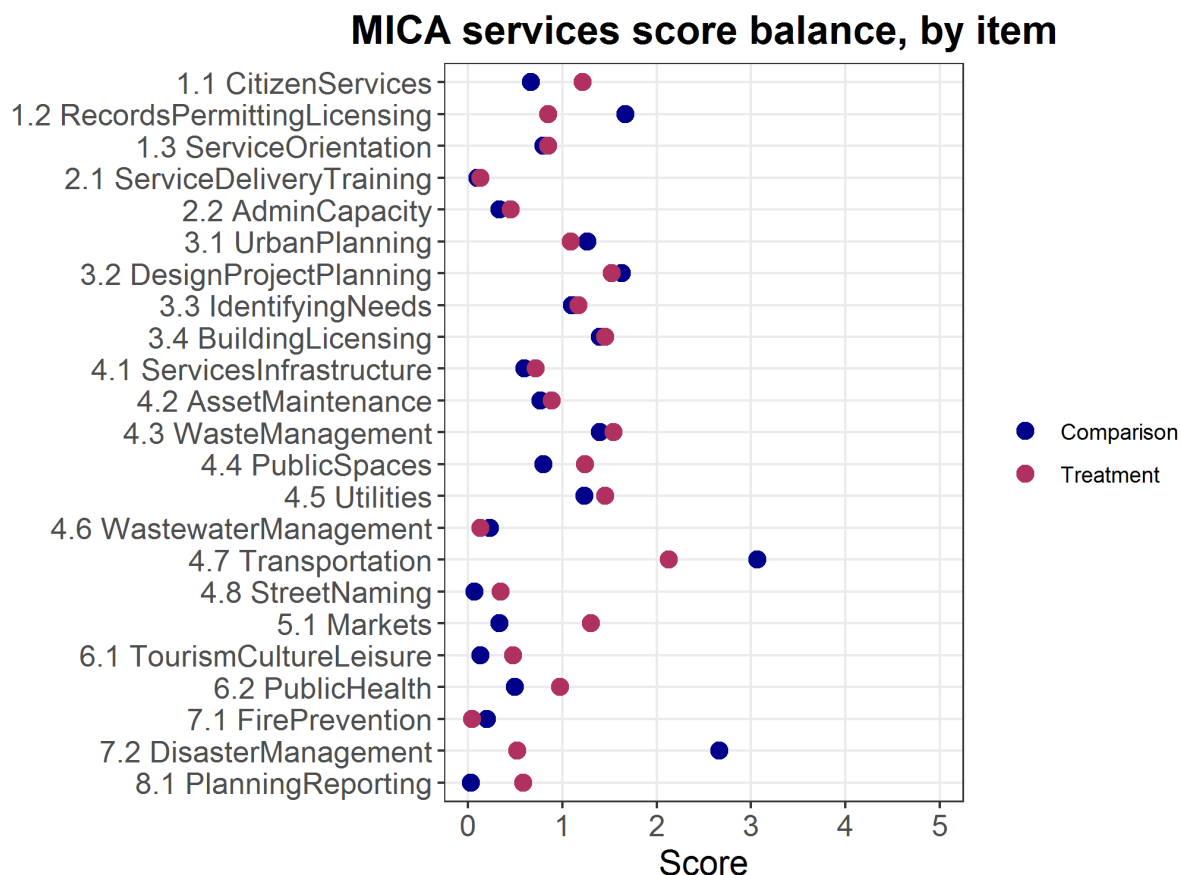
For GESI scores, the pattern is the same and balance metrics are all improved due to a slightly higher overall GESI score in the comparison municipalities. Differences remain high, however, and any impact estimates using these figures will still require scrutiny.

ITEM-LEVEL BALANCE

Unfortunately, small sample sizes prevent examination of individual municipalities or individual MICA measures. Still, examining item-level variability and balance can offer insight into the properties of the evaluation data and highlight areas where inference may be more or less credible.

As noted above and shown in the figure below, there were some differences between the CITIES and comparison municipalities across several MICA services items. The treatment and comparison municipalities' scores differed across overall MICA services items by around 0.1 on average. MICA items related to training in service delivery (2.1), urban planning (3.1), capital projects (3.2), and maintenance of assets (4.2) were particularly similar across assignment groups. The graph below shows the average MICA services scores across the 23-matched treatment and comparison municipalities for each item. Overall, the 15 comparison municipalities scored slightly higher, on average, than the 15 CITIES municipalities.

FIGURE 14: AVERAGE MICA SERVICES SCORES BY ITEM AND ASSIGNMENT (FINAL ET MICA SCORES)



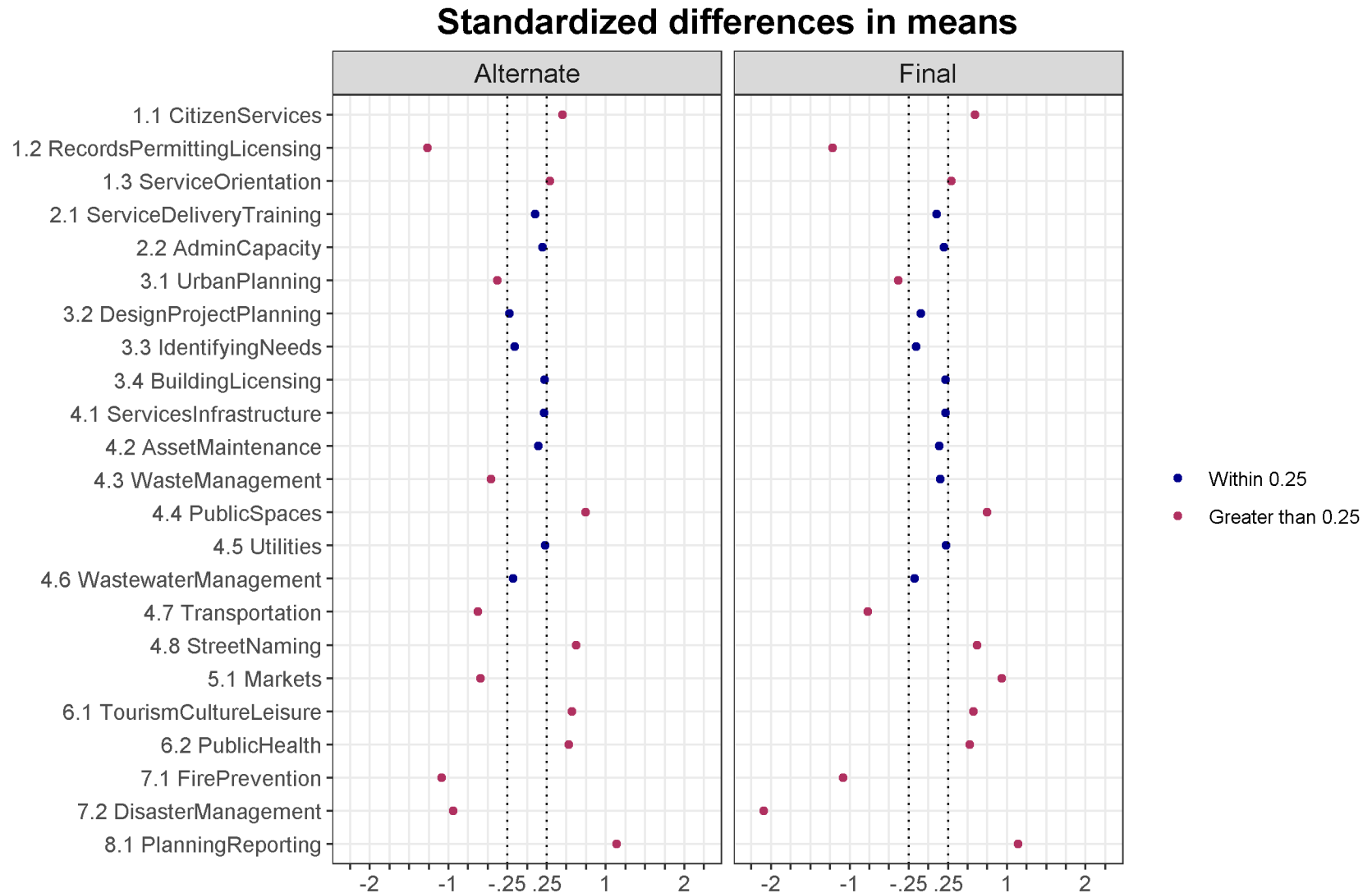
The alternate scoring shows the same general pattern, but with reduced variability across items.

While the above graphic is descriptive, there are specific statistical tests to determine baseline equivalency. Following Imbens and Rubin, the evaluation team used a normalized differences approach to assess balance between assignment groups.²² This method calculates a statistic based on the difference between the treatment and comparison group means, divided by the standard deviation of the data. An absolute value greater than one for this statistic raises concerns, while an absolute value of 0.25 or less indicates particularly strong balance. Normalized differences also help assess whether any potential imbalance can be addressed in the analysis phase.

The table below shows the results of the normalized differences for the 23 MICA services items across both sets of MICA scores.

²² Guido Imbens and Donald B. Rubin. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge: Cambridge University Press, 2015.

FIGURE 15: STANDARDIZED MEAN DIFFERENCE OF MICA SERVICES SCORES, FINAL ET AND ALTERNATE MICA SCORES



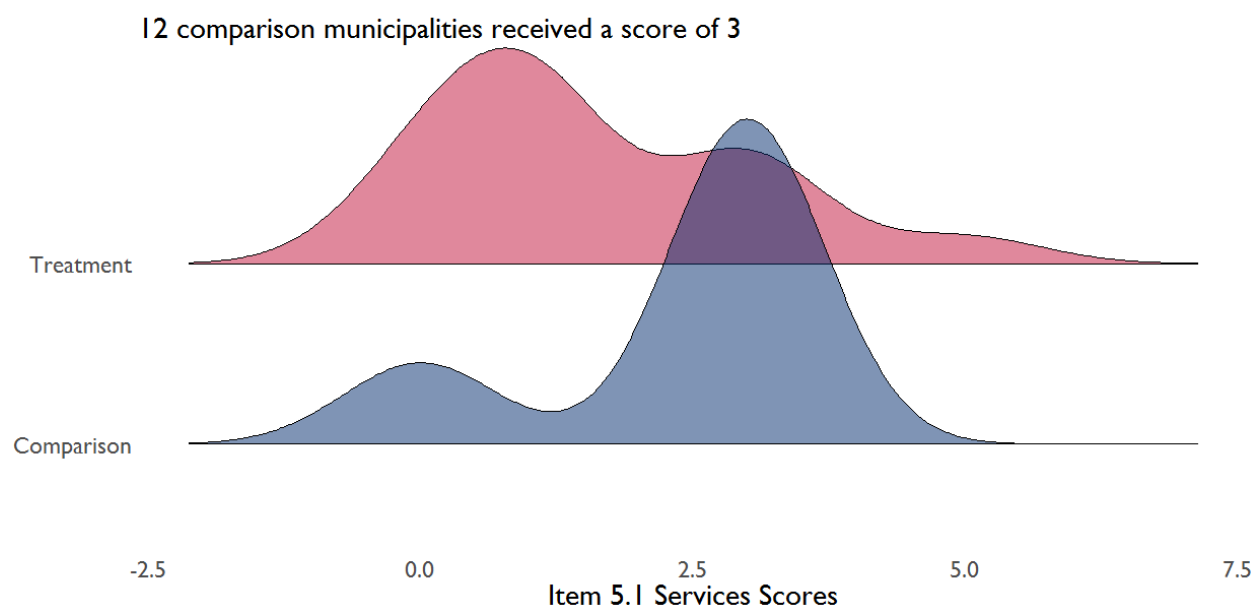
The alternate scoring actually has an additional item whose standardized mean difference lies outside the desired range of 0.25. However, the magnitude of the differences is consistently lower in the alternate scoring, contributing to better overall balance and a closer match in overall scores.

Note that sections 5-8 have larger differences across treatment and control, an issue that is only slightly mitigated in the alternate scoring. On the other hand, only one of the section 3 MICA services items had a normalized difference beyond the threshold of 0.25. This may suggest that, overall, both assignment groups are similar in their ability to plan and identify needs.

It is important to bear in mind that the overall MICA score is the outcome of interest, rather than change in any single item. Although there may be imbalance on many of the items above, the aggregate balance for all items and across all municipalities is of primary concern. The evaluation team will use the information in item-level variability to understand the properties of the MICA scoring and, potentially, develop appropriate weighting schemes to ensure that inferential comparison across assignment groups is balanced. See Annex 4 for the full balance table by item.

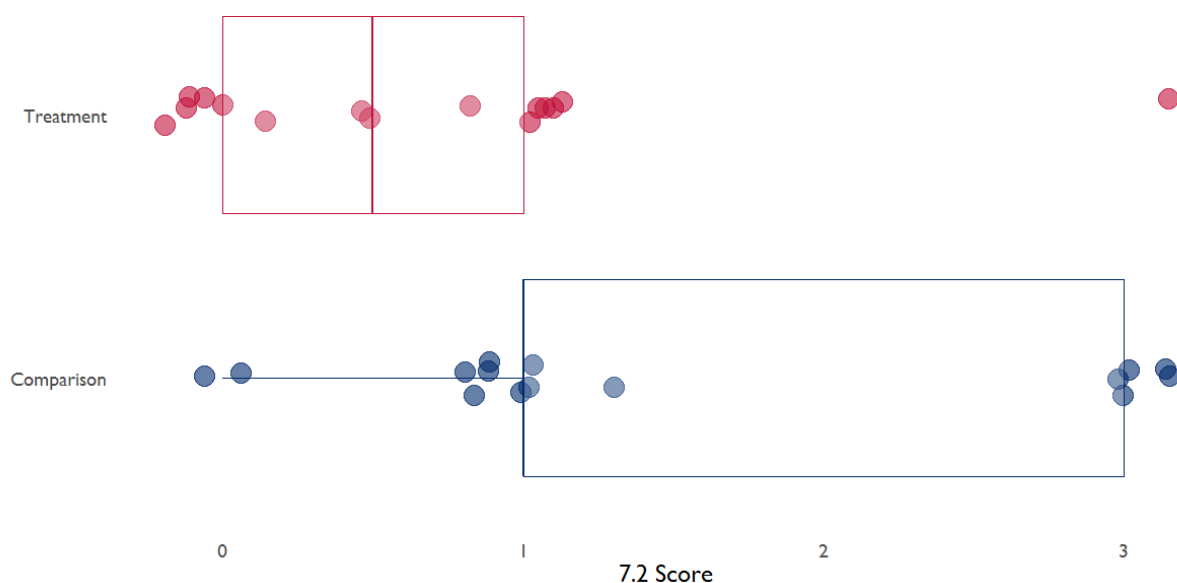
Despite similar scores across a few of the assessment categories, there were MICA items for which the gap was much larger or where one assignment group received a larger share of high scores. For example, CITIES's 15 municipalities received an average score of 1.6 for services item 5.1, which covers a municipality's capacity to inspect markets, while comparison municipalities received an average of score 2.4. This difference was largely driven by the fact that one treatment municipality, Dair Alla, received a score of 5, defined as, "The municipality actively and systematically carries out inspection of public markets, defining the goods that may be traded and goods that may not be traded there", and four municipalities received scores of 3, which has the same definition as a score of 5, but with inspections happening only occasionally. As shown in the graph below, the distribution of scores for item 5.1 varied in its distribution between the treatment and comparison groups.

FIGURE 16: ITEM 5.1 SCORE DISTRIBUTION BY ASSIGNMENT (ALTERNATE MICA SCORES)



Although a third of the treatment group received a score of three or higher for item 5.1, the distribution of scores was wider compared to that of the comparison group, which only had three municipalities that did not receive a score of three. This was not the only case where the evaluation team's scoring was significantly higher than that of CITIES. This was notable for item 7.2, which covers calamity management. The CITIES team gave six of the 15 matched municipalities a score of zero for item 7.2. In contrast, the evaluation team only gave two municipalities a score of zero and five municipalities a score of three, pushing the average item-level score for 7.2 to 1.57. Discussions with CITIES revealed that they only provided scores above 0 for 7.2 if the calamity plan was of “high quality,” a subjective criterion that may explain some of the difference between assignment groups.

FIGURE 17: ITEM 7.2 SCORE DISTRIBUTION BY ASSIGNMENT (ALTERNATE MICA SCORES)



The large differences shown for item 7.2 appears to be the only example where there was a clear and substantive difference in applying the scoring rubric between CITIES and the evaluation team. The alternate set of scores goes far to reduce the scoring differences to a more tolerable level (see Figure 15 above), but does not fully eliminate the differences. In order to help reduce or eliminate these instances in future, the ET recommends closer coordination between the CITIES and ET MICA teams during the endline data collection. This could include ET members accompanying CITIES interviewers when doing MICAs at endline, and vice versa. The ET also expects to develop additional probing questions during the endline MICA study to identify the factors that help explain areas institutional capacity growth as measured by MICA, and also as measured by a variety of other diagnostic tools that CITIES has developed.

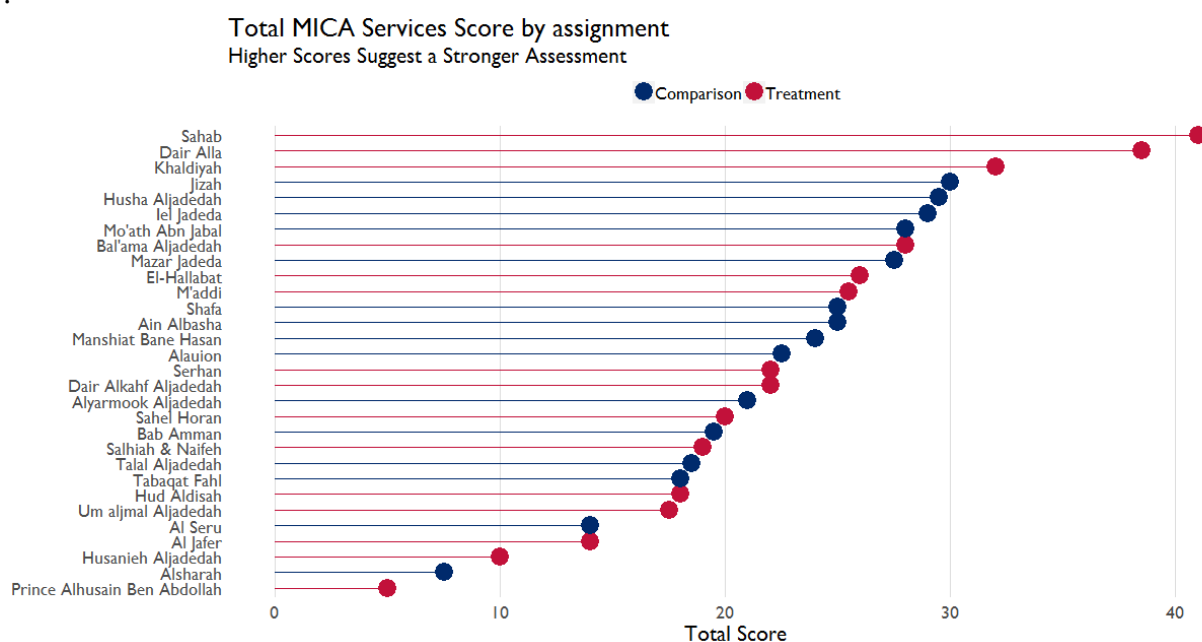
Item-level analysis across assignment groups generally shows that the ET used interim scoring more often; for example, giving the score of 0.5 46 times compared to 19 times by CITIES. The ET explored the use of half-scores with the CITIES team to understand when to best utilize these since the MICA tool has clear integer categories. Additional discussion with CITIES staff may be warranted between this round of data collection and the next implementation of the MICA in the comparison municipalities to ensure that the ET is properly aligned with the MICA methodology. The item analysis findings suggest that there are

gaps between assignment groups on many of the assessment criteria. This is to be expected to some extent given the purposive selection of CITIES municipalities.

MICA Services Scores by Municipality and Region

Reviewing the MICA scores by municipality shows that the CITIES municipalities reached a higher total MICA services score. The highest possible services score is 115. The highest scores in the 15 selected treatment and 15 selected comparison municipalities were 41 and 30, respectively, with a median total score in the CITIES municipalities of 22 and a median score in the comparison municipalities of 24. These close median values can be seen in the figure below, which shows a large grouping of total scores between 20 and 30.

FIGURE 18: TOTAL MICA SCORES BY ASSIGNMENT (ALTERNATE MICA SCORES)



At the regional level, the item-level scores were somewhat similar in the North region, where there are eight CITIES municipalities and 10 comparison municipalities, with average item score of just under 1 across assignment groups.

TABLE 9: MICA ITEM SCORING BY REGION (ALTERNATE MICA SCORES)

Region	Comparison			Treatment		
	Mean Item Score	Item SD	Max Item Score	Mean Item Score	Item SD	Max Item Score
Central	1.05	0.90	3.00	1.43	1.37	5.00
North	0.92	1.12	5.00	0.90	0.92	4.00
South	0.80	1.07	5.00	0.61	0.77	3.00

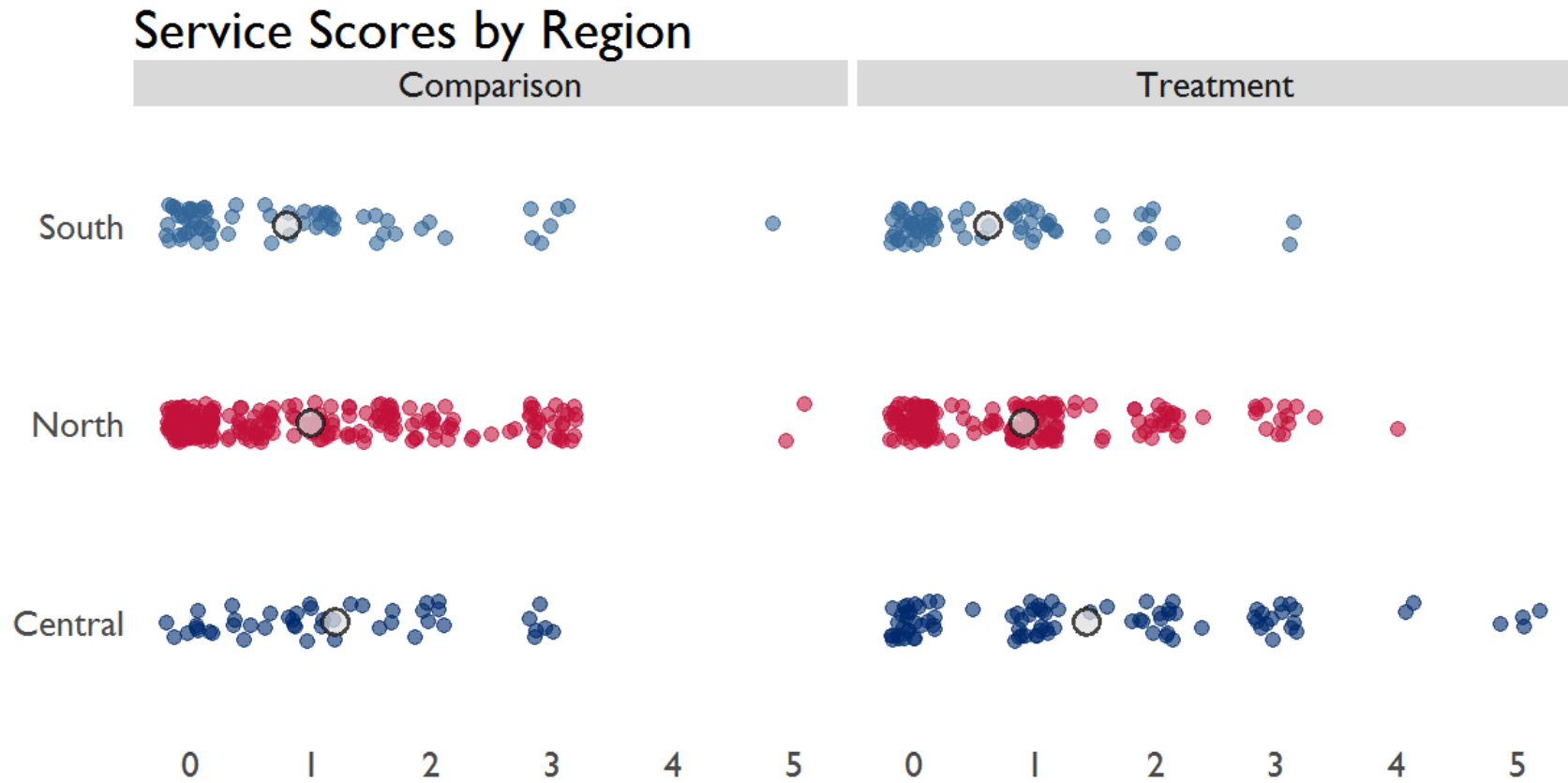
Both comparison group central region municipalities had the highest number of employees within the comparison group and the second (Ain Albasha) and third (Jizah) highest number of employees overall. These are factors that are relevant for MICA scoring, e.g. availability of staff, although the exact relationship between contextual factors and MICA scores will need to be explored through additional inferential analysis using the population survey data.²³ The ET found significant variance and reliability issues with contextual indicators collected through the MICA, such as salaries and budget expenditure, suggesting that these values should be confirmed with national ministries during future data collection rounds.

The relatively low average item score, relatively high item standard deviation, and maximum score of 5 in the South region for the comparison group suggests a somewhat wider range of scoring frequency compared to the CITIES municipalities in the region. Indeed, less than half of the scores in the South region were a 0, with a score of 1 comprising 20 percent of all comparison item scores in the region, and a score of 3 given to 10 percent of comparison region MICA items. In contrast, 51 percent of the CITIES items in the South region received a score of 0 and a score of 3 was only given twice across all MICA services items in the region.

Within the comparison group, the services scoring at the regional level has a similar distribution, with item-level scores generally skewed toward 0. Of the two comparison municipalities in the central region, neither received a score above 3, as shown in the figure below. The distribution of item scores in the figure also shows that the North region were somewhat similarly distributed across assignment groups, albeit with a higher density of items receiving a score of 1 in the CITIES municipalities; 35 percent of CITIES items received a score of 1 compared to 13 percent of comparison items in the North.

²³ There is a 0.57 Pearson correlation coefficient between employees and total MICA score ($p < 0.01$), but this is likely not robust to multivariate analysis given the many other confounders for municipality performance. Budget expenditure and own revenue have negative correlation coefficients -0.27 and -0.12, respectively, but these are not significant.

FIGURE 19: AVERAGE SERVICE SCORES BY REGION



GESI Scores

Although at the aggregate level there is little equivalency on GESI items, at an item level they are paradoxically similar across the CITIES and comparison municipalities. The improved equivalence at the item level is largely a function of the GESI measures themselves. This binary, 1/0 scoring metric was designed to capture the presence of inclusive practices within municipalities. The minimum score will always be a zero and the difference between scores is limited in magnitude. As shown in the table below, the comparison municipalities did not have relevant GESI items for 14 of the assessment criteria. There was a difference in average GESI scores for nine of the 17 GESI items between the CITIES and comparison municipalities. In general, the normalized mean difference for the GESI items suggests that there are differences across assignment groups, with only one item, 3.3, falling below the absolute value threshold of 0.25 mentioned above.

TABLE 10: GESI ITEM BALANCE STATISTICS

GESI Item	Treatment Mean	Comparison Mean	Standardized Mean Difference
1.1 Citizen Services	0.07	0	0.07
1.3 Service Orientation	0.07	0.2	-0.13
2.1 Service Delivery Training	0	0	0
2.2 Admin Capacity	0.13	0	0.13
3.1 Urban Planning	0	0	0
3.2 Design Project Planning	0.4	0.13	0.27
3.3 Identifying Needs	0.27	0.2	0.07
3.4 Building Licensing	0	0	0
4.1 Services Infrastructure	0.2	0	0.2
4.3 Waste Management	0	0	0
4.4 Public Spaces	0	0	0
4.7 Transportation	0.27	0	0.27
5.1 Markets	0	0.2	-0.2
6.1 Tourism Culture Leisure	0.07	0	0.07
6.2 Public Health	0	0	0

GESI Item	Treatment Mean	Comparison Mean	Standardized Mean Difference
7.2 Disaster Management	0	0	0
8.1 Planning Reporting	0.07	0	0.07

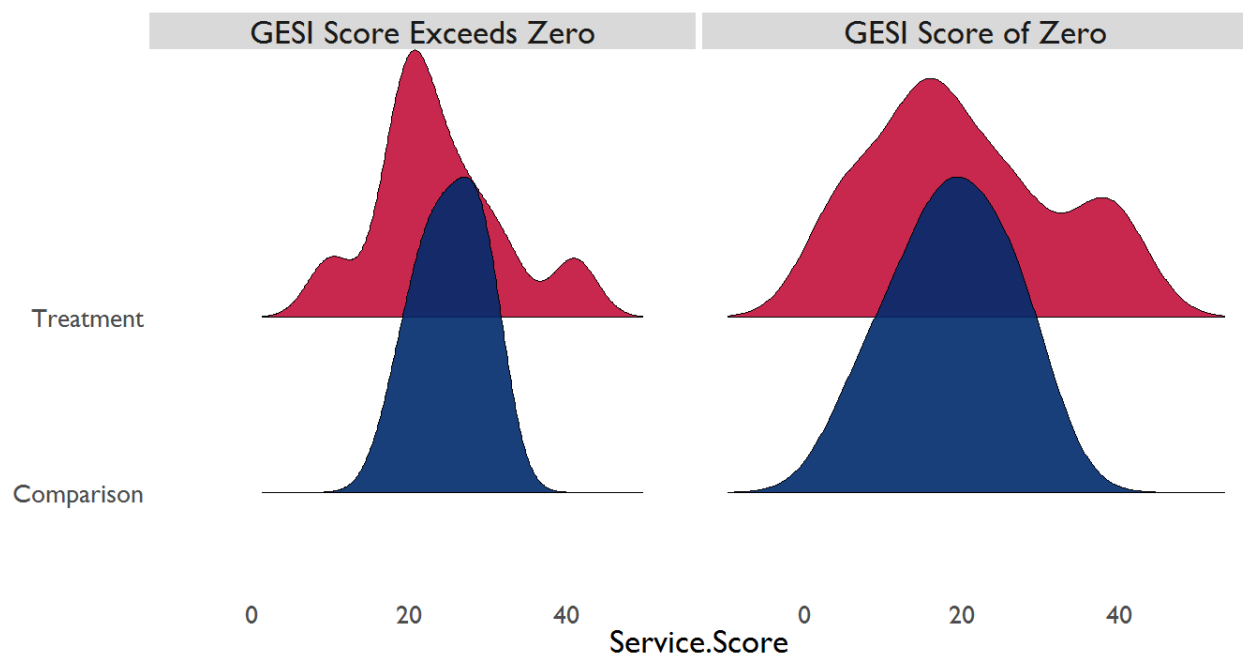
Looking at the items where comparison municipalities did receive GESI scores, the following municipalities had GESI items:

TABLE 11: COMPARISON MUNICIPALITIES WITH GESI ITEMS PRESENT

Municipality	Region	Total GESI Score
Ain Albasha	Central	1
Alauion	North	1
Alyarmook Aljadedah	North	1
Husha Aljadedah	North	1
Iel Jadedah	South	1
Jizah	Central	1
Manshiat Bani Hasan	North	1
Mazar Jadedah	North	1
Talal Aljadedah	South	3

Given the low number of comparison municipalities, it is worth disaggregating the data further to see what, if any, relationship there is between GESI item presence and service item scoring. It may be the case that a municipality's ability to provide GESI-related services speaks to overall engagement with citizen needs. For comparison municipalities that had GESI items present, i.e. scored at least a 1 on one of the GESI dimensions, the median services score was slightly higher than that of comparison municipalities that did not have any GESI items present, with a median total services score of 25.5 to 17.5, respectively. The relationship between the GESI and services scores may warrant additional analysis and qualitative study. The top services score in the comparison municipalities had a total GESI score of 7, but the second highest services score had no GESI items. Indeed, one of the comparison municipalities that received a 1 for GESI item 1.3, which is scored based on whether or not there is a mechanism to collect complaints from GESI groups, had the penultimate lowest total score in the comparison group. As shown in the figure below, the distribution of total scores between municipalities with and without GESI items has a similar skew.

FIGURE 20: DISTRIBUTION OF MICA SCORES BY GESI PRESENCE AND ASSIGNMENT



Given the low number of overall GESI observations, tracking changes in the GESI items may serve as a way to track the types of interventions municipalities receive and the use of, for example, small grants in the treatment areas. These data may also be helpful when combined with the general population survey for additional inferential analysis to see what the relationship is between GESI-related services and overall perceptions of service delivery. Given the range of service scores and limited GESI scores, it is not clear that there is strong relationship between the two based on the current MICA assessments.

There are two key conclusions from this assessment:

- At an aggregate level, the MICA services scores for the 30 matched municipalities is within a standard threshold for equivalency.
- Item level analysis is important for understanding the drivers of difference between assignment groups, but given the nature of the MICA (limited range of scores, inherent rater error, etc.) and the fact that ultimate outcome of interest is the change total MICA scores, any item-level imbalance is important to note, but not detrimental to the evaluation design.

IMPLEMENTATION FIDELITY

As a result of coordination with the CITIES team, there were several updates to implementation that were not planned at the time of the evaluation design proposal. As part of its outreach efforts, CITIES implemented a nationwide public awareness campaign. From March through mid-June 2018, CITIES worked in all 99 municipalities to share information on decentralization with key stakeholders. This included outreach to local community members, local councils, municipal councils, and mayors to address the decentralization and municipalities laws. These activities may have implications for the evaluation in that they may affect certain MICA items and qualitative responses during future data collection rounds since citizens may be more aware of the role of their local government within the context of

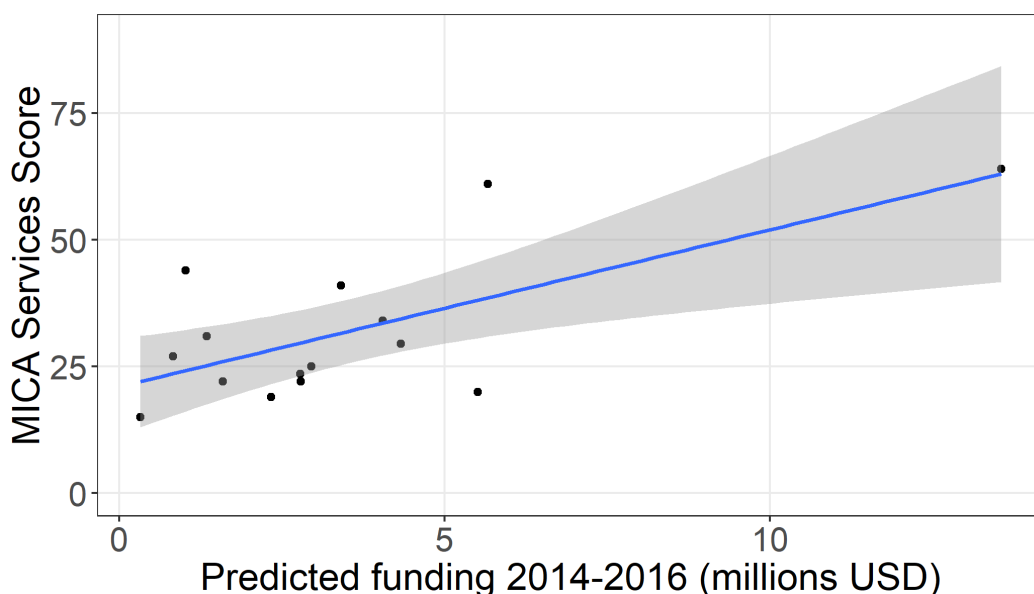
decentralization. One concern that was raised was whether this would undermine the validity of comparing treatment and control municipalities. Given that both treatment and comparison groups received the outreach on decentralization, the evaluation team would expect that, on average, equivalency would be maintained since targeting was universal. In its inferential analysis, the evaluation team may need to investigate the relationship between these activities, covariates, and the outcomes of interest.

Another point of interest will be the rollout of grant activities in CITIES Year 3 (FY 19). Funding to support improvements in municipal operations and service delivery will be the primary catalyst to realize any shift in public perception about municipal responsiveness and effectiveness. Both the timing and objectives of grant activities (municipal operations, service delivery, and NGO support) will be primary determinants of any treatment effect and its magnitude. In this case, activity tracking will help generate learning about what combination of grant activities may be associated with higher or lower treatment effects.

Annex I contains summary meeting notes from the two rounds of implementation fidelity conducted to date, covering FY18-Q2 and FY18-Q3. The evaluation team will continue to monitor CITIES implementation to determine what, if any, changes may need to occur for future analyses.

Another important aspect of implementation monitoring is that of other donor programming in the democracy and governance space. The ET currently tracks three donor activities that may have a direct effect on CITIES indicators, the largest of which is a World Bank grant activity for municipalities to cope with the strain on public service delivery due to the influx of refugees over the last several years. The following graph shows the relationship between World Bank funding levels and MICA scores, after controlling for other factors such as municipality population, poverty rate, and refugee share of the population.

FIGURE 21 WORLD BANK MUNICIPAL GRANTS AND MICA CAPACITY SCORES



The simple analytical model predicts that a 1-point increase in a MICA services score attracted an additional \$124,000 in donor grants over the past three years.

This relationship raises two questions: a) whether higher municipal capacity attracts more donor investment; and b) whether there may be some prior causal effect of donor funding causing an increase in MICA scores prior to or contemporaneous with the CITIES treatment. In either case, it is critical to incorporate measures of other donor support in order to isolate treatment effects by donor activity and identify whether there may be any interactions among donor activities that have an amplification or degrading effect.

PUBLIC PERCEPTION INDICATORS - GENERAL POPULATION SURVEY

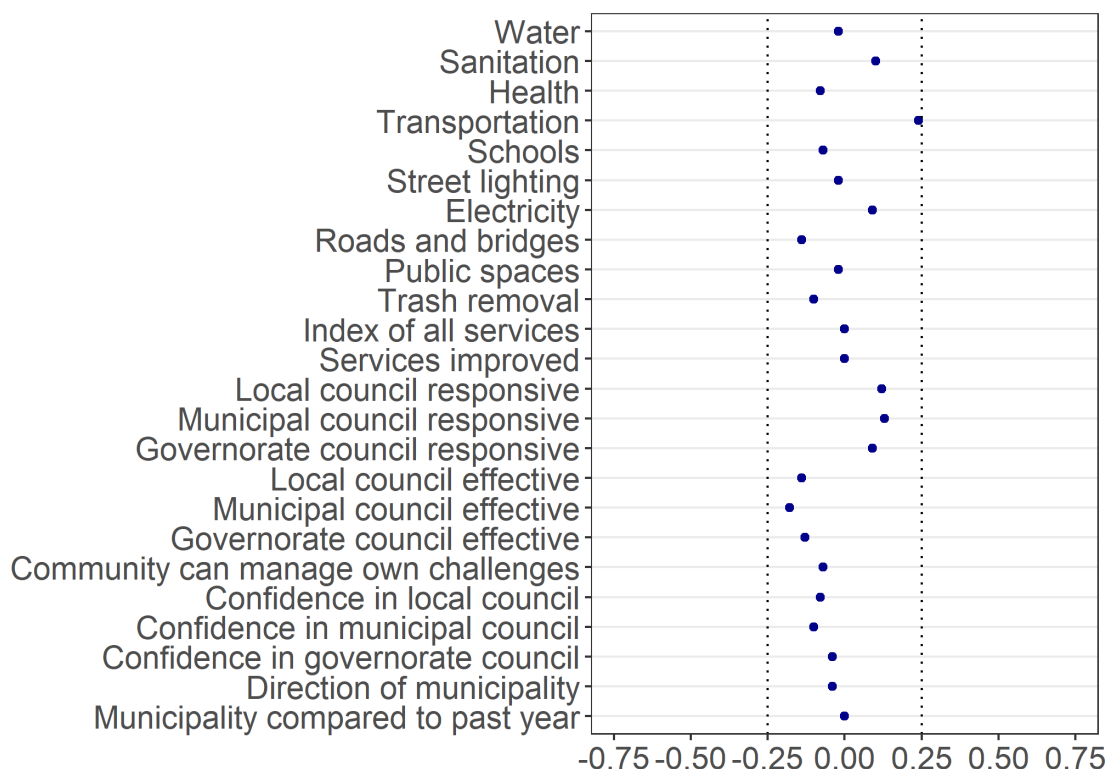
While the evaluation emphasizes the municipal capacity measures (Hypothesis 1) as sufficiently within CITIES' manageable interests to effect change, the evaluation also measures more ambitious indicators based on public perception (Hypotheses 2-7). A survey of nearly 12,000 households was fielded June-September 2018 and included both primary and secondary data collection for the CITIES evaluation. Primary data were direct measurements of CITIES impact indicators such as citizen perception of municipal responsiveness and effectiveness. Secondary data consisted of contextual questions in the local governance and decentralization space, in order to inform a learning agenda examining the broader environment in which CITIES is implemented.

The survey selected 54 municipalities across all 12 governorates with probability proportionate to a municipality's relative share of the governorate population. Of the 54 sampled municipalities, 16 were CITIES treatment municipalities. This led to an aggregate sample of 7,371 respondents in comparison municipalities and 4,592 respondents in treatment municipalities. Sample sizes range widely based on municipality population, ranging from 55 respondents in the Class D municipality of Qatar & Rahmah to 982 respondents in the Class A municipality of Zarqa. To view a detailed breakdown of the sample size by municipality and CITIES treatment status, please see Annex 5.

The broad set of indicators measuring Hypotheses 2-7 are quality of services, municipal responsiveness, municipal effectiveness, confidence in municipal officials, community cohesion, and overall stability. A total set of 25 items measure these indicators across different dimensions. The quality of services indicator disaggregates by type of service, while the responsiveness, effectiveness, and confidence indicators disaggregate by level of government – local, municipal, or governorate. Measures at the governorate level are for contextual purposes only, while measures at the level of local council will have only partial attribution to CITIES. Only the municipal level disaggregation are direct measures of CITIES.

As with balance tests of the MICA data, the following graphic shows the normalized differences of these items from the population survey across treatment and comparison municipalities, where standardized differences greater than 0.25 indicate a potential validity threat to the evaluation measures.

FIGURE 22 STANDARDIZED MEAN DIFFERENCES, CITIES PUBLIC PERCEPTION INDICATORS



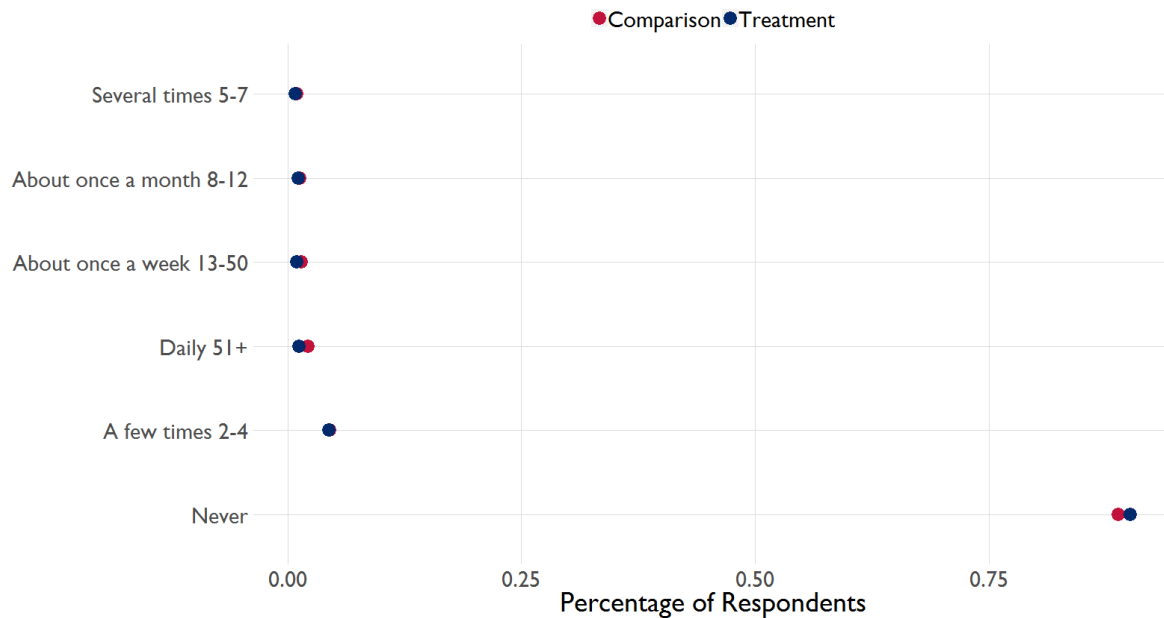
All measures are less than the benchmark of 0.25, indicating good balance on these measures. For additional details on the standardized differences among CITIES public perception indicators, see Annex 5.

CITIZEN ENGAGEMENT AND PARTICIPATION INDICATORS

The use of the MICA captures municipal capacity metrics from the perspective of the municipality stakeholders and through observed characteristics. It is also worth noting how this may or may not translate into perceived capacity from citizens in each municipality. As noted above, the use of the 12,000-person survey sought to measure perceptions across a variety of sectors, including perceptions and citizen engagement. This section provides a brief summary of baseline characteristics within CITIES municipalities. While the focus for equivalency is on the MICA items (as the main outcome of interest), it is worth to highlight where there may be differences across average reported engagement and effectiveness indicators.

The measures reported below are most relevant to providing context to Hypotheses 3, 4, and 6, and fit within the theory of change to provide baseline measures of how citizen engagement may or may not connect to citizen actions and eventually strengthened cohesion. For example, in conversations with CITIES staff, an effort to support municipality websites and engagement through social media was mentioned. As shown below in Figure 1, 88 and 90 percent of comparison and treatment municipality respondents, respectively, reported never visiting their local municipality website.

FIGURE 23: AVERAGE MUNICIPALITY WEBSITE VISIT FREQUENCY ACROSS ASSIGNMENT GROUPS



Other citizen awareness and engagement measures provide similar findings within the context of the theory of change and Hypothesis 3. As mentioned in the background section, CITIES is being implemented during a period of decentralization. Based on the household survey data, around 65 percent of respondents in comparison municipalities voted in the recent elections of 2017, while 49 percent voted in the CITIES municipalities. However, engagement with the electoral process is not necessarily the best measure of citizen engagement as it related to the theory of change.

The table below provides aggregated measures of citizen awareness and engagement on key population survey indicators. Based on the underlying theory of CITIES implementation, the role of citizen action, for example, volunteering or joining a community group, may eventually strengthen community cohesion and make citizens more confident in their municipality. It is notable that the rate of volunteering and group membership were similar across both the treatment and comparison groups.

TABLE 12: CITIZEN ENGAGEMENT AND COHESION GENERAL POPULATION SURVEY MEASURES BY CITIES ASSIGNMENT GROUP

Hypothesis	Indicator	Measure	Treat- ment	Comparison	Difference	Standard- ized Difference
3	Perceived municipal responsiveness	Decentralization awareness	0.64	0.74	-0.10	0.22
5	Ability to address common problems	Community group membership	0.12	0.12	0.00	0.00

Hypothesis	Indicator	Measure	Treat- ment	Comparison	Difference	Standard- ized Difference
5	Ability address common problems	to Contacted or a visited government official	0.13	0.16	-0.03	0.10
5	Ability address common problems	to Municipality meeting attendance	0.03	0.05	-0.02	0.14
5	Ability address common problems	to Volunteered in the community	0.37	0.35	0.02	0.04

There is little statistical difference across both the treatment and comparison groups on contextual indicators related to hypotheses three and five. The largest difference between the two groups on the decentralization awareness measure is still within a general threshold for equivalency. At the municipality-level, Mazar Jadedda has one of the lowest rates of awareness at 58 percent. In contrast, 88 percent of respondents in Al Tafeilah Alkubrah reported awareness of Jordan's decentralization agenda. Also notable is how low municipality meeting attendance was for both groups at three and five percent between treatment and comparison groups, respectively.

AWARENESS AND KNOWLEDGE OF DECENTRALIZATION

One of the learning agenda questions pertaining to democracy and governance asks whether Jordanians are aware of the Kingdom's current decentralization agenda and knowledgeable about the respective government functions and authorities of different levels of government. It is encouraging that 72 percent of respondents reported being aware of the new decentralization and municipalities laws published in 2015. To assess knowledge of actual content of these laws, the survey instrument module on citizen participation includes a set of queries on government functions and revenue sources.

TABLE 13: GOVERNMENT FUNCTIONS QUERIED IN THE HOUSEHOLD SURVEY (SURVEY ITEM D15)

Government function	Primary responsibility	Citation (2015 Municipalities Law)
Maintenance of local assets	Municipality	Article 5.A.5: Cooperate with the competent authorities with regard to the establishment of schools and places of worship, identify their locations, maintain them and manage their activities.
Trash collection	Municipality	Article 5.A.19: Recycle waste, process it, destroy it and determine fees.

Government function	Primary responsibility	Citation (2015 Municipalities Law)
Name streets	Municipality	Article 5.A.4: Opening streets, cancelling and amending them, determining their length and width, pave them, create sidewalks for them, maintain, clean and light them, name or number them , number their buildings, beautify them, plant them with trees, prevent any encroachments on them, monitor the open lands around the streets and mandate their owners to build fences around them.
Providing electricity	National ministry	5.A.7: Coordinate with the competent authorities in the management of providing the population with electricity and gas, and participate in identifying sites of electricity transfer stations.
Numbering buildings	Municipality	Article 5.A.4: Opening streets, cancelling and amending them, determining their length and width, pave them, create sidewalks for them, maintain, clean and light them, name or number them, number their buildings , beautify them, plant them with trees, prevent any encroachments on them, monitor the open lands around the streets and mandate their owners to build fences around them.
Drafting local development plans	Municipality	Article 5.A.1: Draft strategic and development plans and draft a guide on the municipality's needs and priorities, and submit them to the executive council.
Delivery of drinking water	National ministry	Article 5.A.6: Coordinate with the competent authorities in the management of water distribution among the population, organize water distribution and prevent the pollution of springs, channels, basins and wells.
Establishing public markets	Municipality	Article 5.A.9: Organize and establish public markets and identify the types of goods sold in each of them or the prohibition of their sale outside them.
Organize public transportation networks	Municipality	Article 5.A.11: Contribute to the development of public transport networks within the boundaries of the municipality. Establish stops for transport vehicles, identify them, organize them, identify their routes, and participate in identifying their fees when necessary within the boundaries of the municipality, taking into account the provisions of the other laws.
Establishing public parks and gardens	Municipality	Article 5.A.13: Establish areas, parks, gardens, restrooms, and areas allocated for swimming.

While government functions are taken directly from the published law on municipalities, authorities to collect and use revenue consults a World Bank 2005 study of the municipal sector as part of a tourism development assessment.

TABLE 14: REVENUE SOURCE COLLECTION AND USE AUTHORITY (SURVEY ITEM D16)

	Revenue source	Collection authority
a.	Income from municipal investment projects	Municipality
b.	Taxes and fees imposed by the municipality	Municipality
c.	Land and Building tax	National government
d.	Building and construction licensing	Municipality
e.	Permitting fees	Municipality
f.	Fuel tax	National government

The learning agenda explores this information programmatically. For the purposes of assessing knowledge and awareness of decentralization and municipal governance functions as contextual information for CITIES, the following table provides balance metrics for survey items D15 (assessing knowledge of government functions) and D16 (assessing knowledge of revenue collection authorities).

Table 15: Treatment and comparison balance across awareness and knowledge of decentralization and local governance

Knowledge type	Item	Correct response	Treatment	Comparison	Difference	Standardized difference
-	Awareness of new laws	-	0.695	0.731	-0.036	0.081
Government functions	Maintenance of local assets	Municipality	0.414	0.431	-0.018	0.036
Government functions	Trash collection	Municipality	0.982	0.980	0.002	0.014
Government functions	Name streets	Municipality	0.820	0.826	-0.007	0.018
Government functions	Providing electricity	National ministry	0.507	0.519	-0.013	0.026
Government functions	Numbering buildings	Municipality	0.684	0.660	0.024	0.051
Government functions	Drafting local development plans	Municipality	0.377	0.391	-0.014	0.029

Knowledge type	Item	Correct response	Treatment	Comparison	Difference	Standardized difference
Government functions	Delivery of drinking water	National ministry	0.463	0.511	-0.048	0.096
Government functions	Establishing public markets	Municipality	0.646	0.588	0.058	0.118
Government functions	Organize public transportation networks	Municipality	0.308	0.315	-0.007	0.015
Government functions	Establishing public parks and gardens	Municipality	0.785	0.744	0.041	0.094
Revenue collection authority	Income from municipal investment projects	Municipality	0.617	0.620	-0.003	0.006
Revenue collection authority	Taxes and fees imposed by the municipality	Municipality	0.667	0.670	-0.003	0.006
Revenue collection authority	Land and Building tax	National government	0.345	0.359	-0.014	0.029
Revenue collection authority	Building and construction licensing	Municipality	0.775	0.758	0.017	0.04
Revenue collection authority	Permitting fees	Municipality	0.742	0.714	0.027	0.06
Revenue collection authority	Fuel tax	National government	0.606	0.618	-0.012	0.025

None of the balance metrics exceed 0.25, indicating good balance between treatment and comparison groups.

The knowledge items were further subjected to analysis deriving from what is known as item response theory, which is typically used to evaluate the validity of test questions in an educational setting. Analysis from item response theory suggests that the four items testing knowledge of which functions are allocated to the federal rather than local level may be problematic in that they fail to properly discriminate between respondents with high or low knowledge in general. One possible interpretation of this finding is that

while most respondents do have good overall understanding of government functions, understanding of the delineation between local and federal functions remains muddled. Such an interpretation would mirror the findings of the qualitative assessment of the status of decentralization, which found that while government officials might know in theory the distribution of functions, their specific roles and responsibilities with regards to the decentralization remained a point of confusion.

For additional context on public perception data from the population survey, see Annex 9.

CONCLUSIONS

EVALUATION QUESTIONS

EQ1: EFFECTIVENESS

Municipalities suffer from significant structural burdens that impair the provision of quality services to citizens. The most pronounced challenges are those related to municipalities' limited financial resources and revenue streams, the encroachment of the central government on their autonomy and independence, and their overall weak institutional capacity.

Various factors could stall the transition and donors' efforts to improve municipal performance. In addition to the dire situation municipalities are in, coordination mechanisms between local government structures and municipalities are weak; the current structure does not delineate a clear mandate for the various layers, including the local, municipal, executive and governorate councils. The municipal organizational structure has not been reconciled with decentralization requirements. The lack of adequate resources and weak institutional capacity, including that of the LDUs, mean that decentralization does not guarantee economic development. It only ensures that some form of community engagement to determine priorities is taking place. Citizens' expectations are high. This will require quick wins, before the process runs out of steam.

Sector ministries operating at the local level are not under the span of control of the governorates or municipalities and have not seen much power devolved to them. Horizontal links between directorates representing sector ministries and municipalities is also weak. Defining and operationalizing coordination mechanisms between the various local governance and municipal levels is challenging and the features of these mechanisms including for decision-making are still opaque. The implementation of the system has revealed gaps in the current legal landscape that require the revision of existing laws and the development of new regulations.

On the other hand, decentralization requires that the central government ensures the provision of adequate financial resources and the will to devolve power. Municipalities still regard themselves as vehicles for the provision of conventional services and need to grow their development role to attract or implement investment projects, considered key to addressing communities' most pronounced needs.

The effectiveness of the CITIES activity will depend on its ability to navigate the change that can reasonably be expected within this new system. How successful CITIES will be in addressing these issues will partly depend on the training rolled out under Component 1 and 2. Of course, the success of training is highly contingent on the overall enabling environment. The provision of small-scale grants should make some

progress of easing this environment, providing visible benefits to citizens, and demonstrating capacity to citizens in a way that no training program can. The CITIES evaluation will track each of these parts to see what succeeds and where and how this translates to MICA scores and overall citizen perception.

EQ2: SUSTAINABILITY

More time will be needed to determine the sustainability of specific CITIES interventions including the financial management and HR reforms, but a clear theme was the lack of human resource decision making authority at the municipality-level and existing capacity constraints may affect the sustainability of intervention benefits. Moreover, MoMA will need to be on board with the corrective actions recommended through CITIES including the adoption of the organizational structures for the MLDUs and the adoption of the job descriptions for the staff.

EQ3: SYNERGY

Various donors are implementing projects to support the municipal sector and the process of decentralization. While donors continue their coordination attempts, the government's ability to take the lead in this effort remains weak. Baseline data suggests that although there has been some initial coordination between CITIES and other donor activities, there is still room for CITIES, other USAID programs, and other donors to develop synergies and determine how best to complement each other's objectives. It is unclear how each donor activity will affect MICA scores at this point, but should be tracked and analyzed after the next round of MICA data collection is implemented.

INTERNAL VALIDITY AND BASELINE EQUIVALENCY

Comparison MICA scores are generally considered internally valid among MESP raters and with CITIES scorers, but some degree of inherent subjectivity remains. Measurements of institutional capacity have intrinsic uncertainty. Slight differences in interpretation can lead to differences in scores, and no amount of training can eliminate this if it is inherent in the recorded measures. The ET has been vigilant in trying to minimize the error that comes from differences in interpretation. This evaluation has used multiple raters, combined results between them, and incorporated the technical review of the CITIES Component I team lead to minimize measurement error across assignment groups. There are some measures, and some municipalities, where error is relatively high. This error presents risks to the impact estimates comparing individual municipalities, but in aggregate the error is tolerable and does not limit the evaluation design.

Acknowledging that the error level is tolerable, there is still a need to be cautious in inferential analysis. The evaluation found a low-to-moderate relationship between the measurement error and the baseline equivalency between assignment groups. This potentially raises questions about the ability of the evaluation to conduct robust inference. The reduction in scoring error reduced this overall measurement error, but it presents the need to be cautious in inferential analysis. To address this concern, the ET concluded that sensitivity analysis of equivalency using multiple configurations across matched, partially matched, and the full CITIES sample are needed to assess the validity of estimates.

There is baseline equivalency across treatment and comparison municipalities at the aggregate level for MICA service scores, but there is a much greater degree of variability at the individual item level. After reviewing two sets of scores across six configurations, the evaluation team found that the distribution of services scores

at the aggregate level are balanced across assignment groups for matched (15-15) and partially matched (15-23) municipalities for MICA service scores, but not sufficiently balanced for GESI scores. This implies that on average the CITIES and comparison municipalities are statistically similar for MICA service scores. At the MICA element/item level, due to a higher degree of item-level variability, inference at this level will require triangulation with other data streams. Overall, our analysis of equivalency provides hope that any measurement of difference at endline can be attributed to CITIES rather than unobserved characteristics or is the result of selection bias.

Estimates of municipal capacity development are best assessed across a range of configurations. The process of harmonizing the MESP raters to a set of final internal scores and then an alternate set of scores based on a review by CITIES' Component I Lead led to a shift in the statistical metrics used to assess baseline equivalency. The final internal scores suggested that one set of treatment and comparison municipalities was particularly well-balanced, while the alternate scores suggested another set of treatment and comparison municipalities was particularly well-balanced. If a choice is to be made between alternate and final scores, the ET recommends proceeding with alternate scores as a basis for midline and endline data collection. However, given these cross-judgments, a sensitivity analysis approach across both sets of scores will be useful to employ. At endline, any genuine and durable treatment effect should persist across both sets of estimates.

IMPLEMENTATION FIDELITY

Careful attention to the schedule of CITIES's activities suggests a robust initiative that has seen some delay in reaching its intended scale for a variety of reasons, some beyond the control of CITIES. The CITIES evaluation has been carefully designed around the schedule of CITIES's inputs and work plan. As occurs with many donor initiatives, there has been some expansion of scope, with stakeholders making requests outside of the original work plan or expressing a desire to spend more.

Since the start of implementation, CITIES has done a rapid education initiative on decentralization, rapid improvement projects to secure dumpsters for large municipalities, a regional development plan across municipalities, and engaged on a pilot in three governorates (Jerash, Madaba and Ma'an) of financial management and decentralization training. We conclude that these activities will not affect the design or implementation of the evaluation as they were either unlikely to have a measurable impact on the outcomes of interest in and of themselves or were provided across assignment groups. That said, it is important to continue to track such activities and to maintain communication with CITIES to determine whether and how these requests affect planned activities and resource allocation.

GENERAL POPULATION PERCEPTIONS

USAID set out a grand vision to promote Jordan's decentralization agenda, one that spans citizen perception of government responsiveness, effectiveness, and confidence in leaders. This is a tall order for an activity that was designed to help municipalities build capacity to develop own revenue, handle the responsibilities of decentralization, and to improve citizen engagement and constituent outreach. In reviewing the schedule of activities and how faithfully they have been implemented, it is clear that donors should not expect a dramatic jump in citizen perception. Citizen perception of government has a complex and non-linear relationship with the actual performance of government institutions. CITIES is positioned

to help municipalities realize changes institutionally and capacity-wise, but it remains to be seen how this will be perceived by constituents.

RECOMMENDATIONS

Strengthening Implementation Effectiveness and Sustainability:

- CITIES should focus more on contextually nuanced and technical training than awareness raising efforts especially when targeting municipal staff and leadership.
- To the extent possible, CITIES should tie grants to needs identified in the needs manual sanctioned by official process.
- CITIES should support the unification of all capacity building efforts under one institutional umbrella whether it is MOMA's training directorate or another institution to increase sustainability and advocate that such efforts are sequenced to ensure effectiveness and knowledge retention and application.

Synergy

- USAID and CITIES should continue, and increasingly take an active role, to strengthen inter-donor coordination related to support for municipal services and decentralization.
- CITIES and the next ET should support efforts to capture learning generated by the multiple donors in the same development space as CITIES, and its application.

Contextual Understanding:

- CITIES and the next ET should track the following to inform understanding of project context:
 - Changes in legislative and regulatory framework;
 - Regulatory, relational and technical dimensions of the various municipal functions;
 - Relational dynamics between municipal leadership staff including mayors, executive directors and local and municipal council members;
 - Automation process in municipalities;
 - Resource allocation to municipalities and governorates and related decision-making processes; and
 - The National Strategy of Decentralization and the overall direction of decentralization.

Implementation Fidelity Tracking:

- CITIES should develop a standardized component-level or activity-wide tracking system for monitoring activities by date, type, and location. This will help improve implementation, measure scope creep, and provide inputs for further internal and external study.

Evaluation Implementation:

- Considering reliability issues with contextual indicators collected through the MICA, such as salaries and budget expenditure, The ET should confirm values with national ministries during future data collection rounds.
- Considering that in comparison municipalities that had GESI items present, the median services score was slightly higher than that of comparison municipalities that did not have any

GESI items present, the ET should look into conducting additional analysis and qualitative study to understand the relationship between the GESI and services scores.

- The ET and CITIES should maintain or even expand the current level of coordination across the implementation and evaluation teams. The ET used the CITIES' MICA tool as a data collection instrument, which required training and trust, and supported the ET to mitigate rater error. To support this, USAID should continue to support the open communication between the ET and IP stakeholders and consider what additional IP tools may be relevant to incorporate and complement existing data.
- The ET and CITIES should consider holding additional discussions before the next round of data collection to ensure the ET is properly aligned with the MICA methodology.
- The ET should generate impact estimates across both "ET Final MICA Scores" and "Alternate MICA Scores." At end line, any genuine and durable treatment effect should persist across both sets of estimates.
- Even though comparison MICA scores are generally considered internally valid among MESP raters and with CITIES scorers, the ET should conduct sensitivity analysis of equivalency using multiple configurations across matched, partially matched, and the full CITIES sample to assess the validity of estimates.
- The ET should continue to track donor activity until the next round of MICA data collection is implemented, and incorporate that data into the impact estimates to help isolate the effect of CITIES from the potential effect of other activities.

ANNEXES

ANNEX I: NOTES ON IMPLEMENTATION FIDELITY

FEBRUARY 2018 MEETINGS

Component I (CI)

MICA: All MICAs were completed and a final report was prepared. Overall conclusions reveal that first class municipalities perform better than municipalities in other categories indicating clear correlation between size and performance. At the same time, Ramtha scored in the early 20s although the municipality should have scored better. Irbid scored highest, followed by Zarqa then Salt. Class A municipalities usually scored in the 30s. Statistical conclusions cannot be drawn. No municipality showed any sign of activity planning.

Another important correlation was evident between geographic position and performance. Remote municipalities performed weaker than those in the center, including Deir Al-Kahf and Salhiyyeh and Nayfeh. Municipalities in the South also performed weaker than those in the Center and North. According to the component lead, this is related to the financial system, and the areas from which various Ministers hail.

The data did not reveal a specific task that is being performed better in class A municipalities than in other municipalities. CI lead said he could not but conclude that all municipal departments are weak. Small municipalities do not have civil engineers or road engineers. Legal staff is also absent. Civil engineers, when they exist, tend to be very weak. They are seconded to their positions. Sometimes municipalities have dedicated positions for engineers.

MICA Design: The discussion covered rubric issues. CI lead said some indicators should have been better calibrated. Under urban planning, the presence of an urban plan and municipal ability to develop plan should have been split. A great deal of effort and time was spent training team members and bringing them into the process. The subjectivity of consultants was nevertheless acknowledged as a factor affecting results. CI lead said the tool is not suitable for thorough assessment, and interventions cannot be based on its results.

There are performance indicators based on MICA but those can be assessed objectively. CI lead said that some scores are expected to improve. If a “very basic system for participatory budgeting” is improved, the score will go up. According to him, CITIES will have the chance to improve a few scores.

Service Delivery Improvement Plans (SDIPs): On the 12th of March 2018, all SDIPs were completed. Priorities include street maintenance, LED lighting, recycling, cleanliness and public services, sewage systems, and transportation. Priority setting for SDIPs is defined by the community and not by MICA. “MICA is more technical. SDIP is about perceptions.”

There are two streams of funding for municipalities, one through the ministries and one through the governorate councils to which JOD 220 million have been allocated for investments.

Grants: CI lead expressed some concern that USAID's procedures for allocating grants may cause further delays. The budget for grants is now \$5 million for all governorates and 33 municipalities. This modest amount, according to CI lead, may also negatively affect the project.

According to him also, CITIES will need to get grants out to municipalities in the next 6 months. If grants are late, the success of CITIES might be jeopardized. CI lead gave the example of the Al-Husseiniyyeh's mayor who told him that CITIES training is very beneficial but that his municipality "needs a loader." The lack of equipment is impeding the ability of the municipalities to perform their job.

Expectations are high. CI lead emphasized that showing up in municipalities is enough to raise expectations.

Component 2 (C2)

Financial Management: A financial diagnostic has been completed in half the treatment municipalities. Thirteen areas were assessed. The scoring of financial operations system was done to gauge compliance levels and identify gaps in financial operations. Areas where corrective actions can be applied were identified including those of organizational structure, job descriptions, and fixed asset register.

The WB and the Spanish as part of their Qudra project are also working on financial management in three of the thirty-three treatment municipalities. The WB conducted an assessment related to IT systems. CITIES is looking into ways to collaborate in this area. C2 is planning to work in Year 3 on developing an "appropriate" financial management system.

IT Systems: C2 lead is keen on maintaining a strong relation with MoMA. MoMA has been working on automating various municipal systems. For one, financial units were linked to the central government. MoMA is rolling out the system in all municipalities. Some municipalities are not adopting the changes because they did not receive enough training.

Human Resources: Visits were completed to 17 municipalities and data was collected on personnel. A statistical report was developed for each municipality, but the reports had not been released at the time of the ET's meeting with the C2 Team. Reports are usually discussed with target municipalities in cooperation with the MEL and GESI teams.

C2 staff's work with LDUs to develop their structure and propose corrective actions was completed during the 2nd quarter of FY2018. CITIES is planning to work on performance evaluations. C2 is also working with LDD at MoMA. CITIES has inquired if municipalities' organizational structure can be altered, but the Ministry was not in favor. For now, CITIES is going with existing positions and will try to instigate changes through job descriptions. CITIES is trying to demonstrate to MoMA that new regulations require new staffing. Component lead recognizes municipal salaries as a challenge. Current compensation packages will limit the ability of municipalities to attract qualified staff.

Capacity Building: CITIES will be providing capacity building support to MLDUs. MoMA is requesting that CITIES trains LDU staff in all 100 municipalities. CITIES was interested in targeting the 33 treatment municipalities only. The training, (completed in May 2018,) was focused on the preparation process for the needs manual based on the requirements contained in CEP's manual. CITIES decision to focus on LDUs is because the needs guide process is led by the units. C2 lead believes CITIES work would have been easier if the Activity started one year before the elections.

In addition to covering the needs guide preparation process, training is expected to also cover the development of annual action plans, strategic planning and local economic and investment development. MoMA is requesting that support is extended to all 100 municipalities. Municipal and local council members will be trained. Training will be in the form of orientation sessions (2700 elected members). Training will include orientation sessions related to decentralization, the preparation of the needs guide and the mandate of each council. Training will also be provided to governorate and executive council members.

Cooperation with other Donors: Through AFD, the GoJ received a US\$ 50 million loan for the municipal sector. CITIES helped MoMA to develop a financial observatory unit at MoMA. CITIES team visited three municipalities and assessed their system and developed a disclosure policy for MoMA. This support included developing a communication mechanism and the training of 100 municipal financial managers.

Component 3 (C3)

CITIES has dropped the cohort system from its approach. C3 lead said the component tends to be somewhat reactive in its own approach. If the municipality is interested, staff will visit it and hold discussions.

C3 is working with the Municipal Community Outreach Groups (MCOGs) to develop their outreach strategies. The MCOGs are headed by the mayor and include the membership of the LDU Media and PR staff and local council heads, youth and women. Local council members are not concerned that they are not all members. Sometimes, the MCOGs do not await CITIEIS teams to offer their support. They start initiatives on their own.

C3 lead said that the municipalities are accepting the component's team and the component is on target in terms of its activities.

C3 prepared a community engagement review report for the 17 municipalities in which C3 conducted an initial assessment/review. The component is working with MCOGs to identify 3-5 tools to be included in each municipality's Municipal Communications Strategy. The strategy will be endorsed by the municipal council. In parallel, the component will be working with municipalities to develop various communication tools such as websites, social media and townhall meetings.

Component 4 (C4)

Changes in Approach: C4 lead said the approach has shifted. Threats to community cohesion (TTCs) are now being considered. In targeted municipalities, C4 team meets with the Mayor, LDU staff and local council members and explains CITIES approach to promoting community cohesion within the community and municipality. If buy-in is secured, the team asks the municipality to hold a community meeting to discuss threats. About 25 community members representing different sectors attend and engage in a brainstorming session about the factors affecting relations in the community. From there, they ask community members to identify factors "that help the community bond, identifying in the process social and religious values, such as nakhweh, that help achieve that. After threats are compiled and clustered, the team meets with the MCOGs to review the findings, prioritize TTCCs, and discuss potential solutions to address them. All the information gathered is then summarized in the Community Cohesion Intervention Matrix.

C4 team, therefore, helps municipalities prioritize the TTCCs that were identified through this process. This is done with the MCOG. Following that, C4 lead meets internally with other component leads to understand what other components are doing and how work can be coordinated.

So far, introductory meetings have been held in 16 municipalities; TTCCs exploration sessions were organized in 11 and prioritization was facilitated in 6. Interventions to address threat have not been identified yet.

Ranking Threats to Community Cohesion: The process has identified common “threats” in communities. These include the “lack of zoning,” which causes friction between community members that have the potential of escalation. This is sometimes due to arbitrary construction. Community members end up dragging disputes to the municipality. C4 will attempt to address some of these threats.

The second common threat identified through this process is garbage collection, including the location of bins.

The third threat is stray dogs.

The lack of parks and public spaces has surfaced as a final common threat.

Non-service related TTCCs include elections and drugs which cause “agitation” in the community.

Gender Equality and Social Inclusion (GESI) (Cross-cutting component)

GESI has an advisory function not an implementation one. That said, GESI lead said her crosscutting component may end up having some implementing role. GESI lead works with all components to ensure that GESI is mainstreamed. When mainstreaming GESI, analysis and documentation will be undertaken to provide component leads with data to help them ensure GESI sensitive interventions. This crosscutting factor is supported by many stand-alone capacity building and technical assistance interventions.

A GESI analysis was not conducted. GESI lead said she was not able to conduct one. She only did a desk review and produced a short report. The thinking at that time was that a separate GESI analysis would be conducted for each municipality. A GESI municipal review was subsequently prepared for each of the 33 municipalities.

GESI Review Process: The GESI review was designed to cover four activities. In the Municipal Institutional Capacity Assessment (MICA) 18 statements were added to evaluate the status of targeted groups and assess the level of services and relation to municipalities. MICA GESI scores are now being finalized.

The second part covers the HR audits. Eight were done so far. A gender analysis of the HR audit will be conducted.

The third part is the baseline survey. GESI lead met with REACH (who conducted the baseline survey for CITIES’ indicators) and incorporated questions in the survey. REACH provided a short report on GESI.

The last part includes data collection from female elected councils and staff about their challenges, experiences, and aspirations. The data is analyzed and incorporated in each municipality’s profile. The generated analysis will mostly feed into Component 3 and 4 only.

In Year 1 when MoI revised the organizational structure of LDD, the ministry added a unit for community outreach to ensure that all stakeholder needs are identified. The GESI lead started working with the new Unit. The staff was brought together to discuss their tasks then were trained on the decentralization needs manual.

The Unit is expected to hold sessions with community groups. Community outreach/equal opportunities representatives will meet with the community to identify needs at the governorate level. They will focus more at sector needs rather than municipal needs. Because of this engagement, the sequence of engagement and needs identification and prioritization is expected to be much easier next year.

AUGUST 2018 MEETINGS

Component I (C1)

Municipal Priorities for Activity Interventions: After C1 team completed the MICAs with municipal staff and reported on results, the staff engaged in identifying about 6 intervention priorities in each municipality. The identification process yielded common issues across all treatment municipalities. The first common priority identified through this process is street maintenance which was mentioned in 30 out of the 33 targeted municipalities. This will be addressed by CITIES through training on maintenance plans, surveys, assessment of technical quality, implementation plans and other related areas. C1 lead felt that outsourcing quality could also be improved.

Street lighting was identified as the 2nd priority issue for municipalities followed by health inspections, one-stop-shop then solar energy. C1 lead discussed the assessment of solid waste management that the project conducted. According to him, there is no occupational health management in any of the municipalities.

SDIPs were based on MICA results, observations during the MICA sessions, and other assessments. According to C1 lead, activity interventions are informed by SDIPs as well as team observations. Both, in turn, inform municipal plans. Some C1 findings are corroborated by C4 findings related to communal stressors.

Training offered by CITIES will be in the form of on-the-job training followed by coaching to monitor and support the application of knowledge. Training will start in late September. Groups will be organized geographically. Each group will include 6 municipalities with two staff members from each municipality.

Street Naming: The minister requested that street naming is included as a priority area. The tender procedure is in process.

Public parks: Public parks will be a joint activity with C4. The first 4 municipalities are currently being selected for implementation. Municipalities will be asked to identify locations. C4 will mobilize an organization to develop sites and will also provide some equipment. CBOs and the local community will also be mobilized to support the rehabilitation process on voluntary basis.

Solar Energy: In terms of solar energy, C1 carried out a study to determine if the project is technically and financially feasible. C1 is proposing that municipalities contract private investors to build the solar plants to feed municipal needs consuming about 15% of budgets (around JOD 150,000.) CITIES' role will include developing feasibility plans and implementation plans and overall facilitation.

Fleet Maintenance: CI hired an expert to visit all municipalities and train staff on fleet maintenance. The expert will train staff on how to do first line maintenance and how to manage car parks.

Solid Waste: The assessment of solid waste that CI conducted identified community cleanliness as a priority need. Based on the assessment, municipalities have defective route planning. Routes are not the most efficient which may lead to municipalities incurring related high costs. Support for route planning was requested by 16 municipalities and will be provided in 18. The project will support municipalities to develop solid waste management plans that will cover proposals to improve garbage collection, job descriptions, analysis of routes among other operational issues. Support may also include small equipment (\$5000 per municipality). In Ma'an, the municipality has already accepted the plan. Health and safety equipment will also be provided in the amount of \$2000 per year per municipality.

Recycling: Several municipalities requested support for recycling initiatives but CITIES will not provide such support except in a few locations, including Maadi in Salt. Support will be in the form of technical advice for sorting at source. In Sahab, where the municipality was also keen on starting a recycling initiative, CITIES will only help develop a SWM plan for the municipality. CI lead emphasized the importance of supporting visible initiatives including public parks. Recycling is not considered sufficiently visible.

Health Inspections of Food Shops and Markets: GAM has a good system for inspecting professional licenses. CITIES has hired someone from GAM who is ready to undertake related training. The training will cover such issues as the current legal framework, how to maintain a database and carry out systematic inspections and when to resort to the police.

One-Stop Shop Service: CI lead expressed some skepticism about this identified priority need and what it means for every municipality. According to him, municipalities will receive different forms of support in this area. Some municipalities simply want a sign. Some want equipment for a new building. In Zarqa, CITIES will provide equipment. In Irbid, support will be provided under C2. In Salt, CITIES is working with Mercy Corps to provide needed assistance.

Local Development Plans (LDP): Only some municipalities have expressed the need for developing LDPs. Decentralization requires municipalities to have both strategic and development plans in addition to the needs guide. The sequence of how these plans are developed is important. During 2018, the development of the needs guides went ahead without being informed by municipal strategic plans because a few municipalities have such plans. CITIES will roll out a training program in all 33 municipalities to cover gaps in this area. The training will be piloted in Sharhabeel in Jordan Valley. All 33 municipalities will be taught how to collect data.

CI also reported that the development of a regional development plan with two other municipalities (Mo'ath bin Jabal and Tabaqet Fahel; both comparison municipalities) is underway.

CITIES trained MoMA staff. The project provided a 6-day ToT training on the development of LDPs. CI also trained LDUs and executive council directors on how to develop governorate development plans. This, according to CI lead, was a success because twelve governorates produced their own plans. Even though some of these plans will require additional effort to enhance their quality, the various governorates now have plans developed independently from MoPIC. Moreover, and according to CI lead, by the end of August 2018, all governorates would have approved their investment plans.

The new decentralization law requires governorates to develop both strategic and development plans. CITIES advocated that the two be combined and support is provided for the production of strategic development plans. These contain vision and implementation plans. The combined plans have now been approved by the executive councils and referred to the governorate councils where they are being discussed. MoI will not need to approve them.

Reporting: Mayors are required by law to submit reports to municipal councils and then to the Minister of Municipal Affairs. CITIES will train mayors on how to report and will develop a standardized template for reporting. At MoMA, staff will be assigned to follow up on these reports. Reports could be used to inform policy.

Master Plans: MoMA requested that CITIES support the development of master plans, which CITIES will do.

Tracking Inputs: CITIES developed supplementary indicators because many of CITIES' interventions are not reflected by the Activity's primary indicators. Beyond indicators, Municipal Action Plans are updated on a monthly basis. Across all components, the plans capture all activities in all the treatment municipalities.

Challenges: MoMA has high expectations for CITIES. The minister has made several requests for support, including for street naming, master plans and regional plans. MoMA expected that CITIES would provide more support in the form of in-kind assistance. According to CI lead, municipalities were also under the impression that CITIES would be providing more equipment. CI lead said CITIES "can achieve a decent result." He expects MICA scores to move by 10%. He expects cancellations of support for LDPs and one stop shops.

On MICA, CI lead said the tool is useful but "rougher and less accurate" than intended. It also does not capture qualitative changes. For example, the quality of transportation such as reliability of bus services is not captured by the tool.

Component 2 (C2)

Financial Management Assessments

- The financial management assessments will be completed by September 2018.
- Thirteen (13) areas of financial management are being assessed by CITIES.
- Assessments in 25 municipalities have already been completed.
- C2 organized 75 days of orientation sessions for local and municipal council members. The sessions covered the role of the councils and procedural aspects. C2 lead said that the 75 days orientation sessions delayed C2 core activities.
- C2 also trained all MLDUs on the procedural manual developed by CEP. CITIES tried to include MoMA staff in the 3-day training conducted with Partners for Good. C2 conducted trainings for 16 MLDUs and 3 GLDUs, in Jarash, Madaba and Ma'an governorates to finalize the needs list. After the needs lists were approved by the municipal councils, C2 worked with GLDUS to help them consolidate the lists.
- C2 also conducted 12 2-day trainings for governorate council members on budgeting and a Fiscal Reform Project developed manual about budget formulation.
- To assess trainings, CITIES conducts pre and post surveys. The last training attempted to assess buy in of financial staff and commitment to adopt corrective actions. In general,

municipal staff are interested and responsive. Educated young managers tend to be less resistant to change.

Human Resources: C2 completed 19 HR audits.

There is a connection between financial scoring and C2 activities. C2 is planning to work on cash flow, revenue tracking, and budget formulation.

Challenges: C2 lead said he expects some challenges with having municipal staff adopt corrective actions without the provision of financial incentives or a mechanism to motivate and ensure buy-in of new interventions.

Another challenge relates to MoMA's slow process. C2 is planning to train financial managers, HR, local council members, and MLDUs. These trainee groups refer to different units at MoMA who are slow to respond and coordinate.

In terms of progress, two activities are derailed: The development of the local council manual of control. C2 lead asked for the activity to be removed because there is no need for a manual. The decentralization unit will develop a manual for all councils that will cover all regulations. Another activity that will be delayed is the training activity for executive managers, due in part to the government's own delay in hiring executive managers.

Component 3 (C3)

C3 Indicators: C3 lead said that C3 is meeting its targets for indicators relating to public forums. The definition of public forum is an event attended by more than 16 community members. C3 reached 70 public forums instead of the targeted 66 because of this year's SDIP meetings. MCOGs are also counted towards "avenues for youth" because youth are represented on the Groups.

C3 lead said that when MCOGs hold meetings with the community, events are similar to townhall meetings. In Jafer, for example, one of the meetings with community members was able to attract 55 people.

The formation of the MCOGs could be counted towards indicator 1.3 on Municipality Serviced Orientation toward Individual Citizens in the MICA tool.

Challenges: These include:

- Municipalities dropping out of the project;
- Weak MCOGs that need to be more closely supported and encouraged to remain active;
- Other donors have committees in municipalities whose work will need to be coordinated and followed up on; and
- The component's ability to capture information relating to its own activities.

Awareness sessions will be held on the subject of budgets. People need to understand the limits of municipal action and the responsibilities they bear as citizens, including the responsibility to pay taxes. Community members should also understand the municipality's financial cycle, who approves it and how it's approved.

Eight municipalities are already sharing their budgets with their communities. More effort is needed for full-fledged participatory budgeting.

Moving forward consultative sessions and capacity building are needed to build communication tools the MCOG already identified including town hall meetings and community members databases. Automated data with report generating capabilities is also needed.

Component 4 (C4)

TTCCs Identification Approach and Progress: The TTCC profiles for 23 municipalities are complete. Salt and Madaba have just been finalized.

This Activity's framework is discussed and set with the mayor and MCOG. This includes one or more meetings with the mayor or staff to explain the approach and decide on when to hold the exploration meeting with the community. About 25-30 community members attend, usually in addition to municipal council or governorate council members who also invite people to join.

During communal sessions, CITIES staff ask community members to visualize a cohesive community, and then ask them to focus on their community and identify factors that would strengthen cohesion. Threats are categorized as either service related or social. Service related threats are divided to municipal services and other public services. Public services cover transportation and education, water and electricity. The social category of threats includes general threats and youth and women related subsets. Community members are also asked to identify stressors on relationships, including that with the municipality.

Through their efforts, C4 staff try to construct an understanding of what the threats in the community are. In Dleil, for example, the municipality complained about challenges with engaging youth. C4 staff try to understand if this is a mere stressor that is service related or a TTCC. Sometimes service-related stressors can morph into TTCCs. According to C4 lead, transportation can escalate and become a threat to community cohesion in a situation where a boy, for example, sees his sister in a private car with a stranger.

C4's approach follows that of CITIES in terms of first identifying and prioritizing issues within communities then deciding how to address them. Threats are compiled and reviewed through MCOGs and are then prioritized.

Results revealed common threats across a number of municipalities, including the lack of public parks and lack of zoning. Common social issues include drugs.

Priorities, however, were difficult to identify. Social issues are easily ranked, but not other threats. Now C4 is working with MCOGs to decide the nature of needed interventions. In February 2018, C4 engaged the community in awareness sessions on the roles of municipal and local council members. Ninety-seven (97) sessions were held, one in each municipality with a total of 4,055 attendants. C2 held awareness sessions on the roles of councils with council member, while C4 held the sessions for community members. Five of these sessions targeted youth. C4 has then resumed its analysis on the threats to community cohesion. To date, C4 has worked with 25 municipalities to identify threats to community cohesion. In 18 municipalities the threats were ranked, and in 14 appropriate interventions were identified.

Interventions will include procurement or in-kind grants for the rehabilitation of parks, soccer fields, multipurpose rooms, and adolescent centers. On the "softer" side of possible activities are awareness

and cleanliness campaigns, the formation of youth groups and the facilitation of meetings with stakeholders, such as cattle owners and municipal staff. C4 interfaces with LDUs to facilitate C4 activities.

Challenges and Changes: C4 raises awareness about the importance of accountability and role of community in decentralization. Challenges include the high level of distrust between administrative layers, and the widespread misconception about roles and responsibilities.

Some municipalities are conservative. This limits the participation of women. Jafar and Hallabat are examples of such communities. Women don't attend sessions held by CITIES.

In terms of knowledge management and how information is documented and reported, CITIES uses attendance sheets and pre and post surveys. Activity tracking is documented in the Municipal Action Plans.

Conflict Management and Mitigation (CMM) was shifted to Year 3.

Meeting with CoP

Challenges: The CoP reported that the Activity (as of August 2018) is not where it should be in terms of outcomes or impact. Assessments have not been completed yet. C2 activities have warded off relationship problems with municipalities. However, both mayors and communities are frustrated. The component leads are also frustrated because they are not able to implement activities yet. The MICA and SDIP processes are taking too long. Because grants have not been disbursed, CITIES is negatively affected. In October 2018, the Activity can make in-kind donations.

Upcoming Assistance: The grants manual is already approved. Grants will be branded as USAID CITIES grants. There is an ongoing conversation with the mayors about how the grants will affect public perception of municipalities. CITIES is trying to assist mayors to be seen as bringing in additional funding to help the municipality and community.

The CoP reported that thirty percent of each grant will be at the discretion of mayors to spend as they see fit. Seventy thousand will be allocated for service delivery and another 70 thousand for community engagement and cohesion. One and a half million will be granted to CSOs to help municipalities move on social cohesion.²⁴

According to the CoP during this meeting, C2 is not tied to grants but can procure services. C2 has a procurement plan (separate from grants) for organizational development. The component's budget is US\$ 600,000. Seven municipalities were selected to receive equipment, printers, and computers.

Coordination with other Donors: CITIES worked with the French to help develop a financial observatory at MoMA. Cooperation between the CITIES and AFD is strong.

²⁴ Since August 2018 the concept for municipal grants (the reference to 30% of the grants) has been set aside with the project grants now focused on in-kind assistance (about 105K JD per municipality) and FAA grants for civil society (about \$1.6M)

ANNEX 2: INVESTIGATING SCORER AGREEMENT

The Municipal Institutional Capacity Assessment (MICA) is CITIES' primary tool for determining a municipality's service delivery capacity. CITIES has described the tool as a way to "start a conversation" around its implementation approach in each of its municipalities. Change in the MICA scores is also the main outcome of interest for this evaluation, with the comparison of scores between CITIES treatment municipalities and the matched control municipalities serving as the main focus of the IE design.²⁵ Given the importance of MICA scoring for both implementation and estimation of CITIES's impact, assessment of the MICA scores themselves is warranted to ensure reliable, minimally biased estimates.

There is concern about measurement consistency with any data collection effort. Two reasonable people can interpret the same question in slightly different ways, which can affect the reliability of the overall measurement endeavor. In the case of the CITIES IE, the use of the MICA as a primary data source for this IE presents two main challenges: between-team reliability and within-team reliability.

The evaluation team's MICA scoring in control municipalities needed to be aligned with that of CITIES' scoring in its treatment municipalities to ensure between-team comparability. The evaluation team and CITIES needed to have a similar understanding of each MICA item and how to apply MICA scores appropriately. To this end, the evaluation team received two training sessions with members of the CITIES team. The evaluation team also shadowed CITIES staff as they applied the MICA in six Mafraq governorate municipalities.²⁶ The goal of training with and shadowing of CITIES staff was to minimize the variability between each group's scores and overcome the challenge of each team implementing the MICA tool differently.

The second challenge MICA scoring presented as part of baseline data collection was within-team, or inter-rater, reliability. Even though the IE team received training on how to employ the MICA tool, there could still be variability within the IE team on how to appropriately score the same municipality on the same measures. Beyond the training and guidance that the evaluation team received from CITIES, the MICA tool itself has guidance on how to apply scores for each item. For example, on MICA measurement 2.1, Technical Trainings in Service Delivery, the MICA tool provides the guidance to apply a score of 1 if, "The municipality occasionally organizes training on technical topics (e.g. maintenance infrastructure, maintenance of vehicles/equipment, SWM), at request [sic]." This seems straight forward, but there are often cases where conditions on the ground are more ambiguous, such as whether staff were actually able to provide services and their qualifications to do so. If there is a lot of variation within the IE team on what constitutes one score over another, assessment of the comparison municipalities could be biased as would comparison to CITIES' treatment municipalities.

Although the evaluation team used the same MICA tool as CITIES, the mode of entry differed. The evaluation team collected MICA data using an electronic Open Data Kit (ODK) survey. ODK is an open-source survey platform that is compatible with most mobile devices and can save data entry progress. The scoring process involves a series of interviews, relies on extensive notes, and, in general, took place over

²⁵ Other measures, such as citizen perception of local government responsiveness and effectiveness, are more impactful. However, improvement in MICA scores are seen as directly within CITIES' manageable interests to achieve, while changes in citizen perception are subject to multiple factors outside of CITIES' control and are also difficult to measure with the necessary precision to attribute to the CITIES activity.

²⁶ The evaluation team thanks George Hartman, CITIES team lead for Component 1 Service Delivery, for his generous donation of time to support the effort, and to allow evaluation team scorers to accompany his own staff.

a few hours, with follow-up calls as needed. The evaluation team developed the ODK-version of the MICA tool to ensure standardization of entry and to develop a central database for scores. Enumerators took notes during the interview and then re-entered these into the ODK form the day after data collection. By limiting room for data entry error and having data uploaded as soon as it has been entered, the evaluation team in Jordan and at MSI in the US could easily check for issues, track data collection progress, and begin analyzing the scores. In addition to using ODK for the comparison counties, the evaluation team also used the ODK MICA tool to enter MICA scores from CITIES. Through this process, the evaluation team has all of the MICA scores in a single, uniformly formatted data set. For the comparison municipalities, the ODK MICA tool allows the evaluation team to see who scored which municipalities, when the scoring was started, when it ended, and output the data in an easy to analyze file type.

ASSESSING RELIABILITY

Before determining how to address inter-rater reliability, the evaluation team needed to know the magnitude of score variability. To assess how harmonized the IE team was on their MICA ratings, two or more team members scored the same municipality for the following 15 municipalities:

TABLE 16: NUMBER OF RATINGS PER COMPARISON MUNICIPALITY

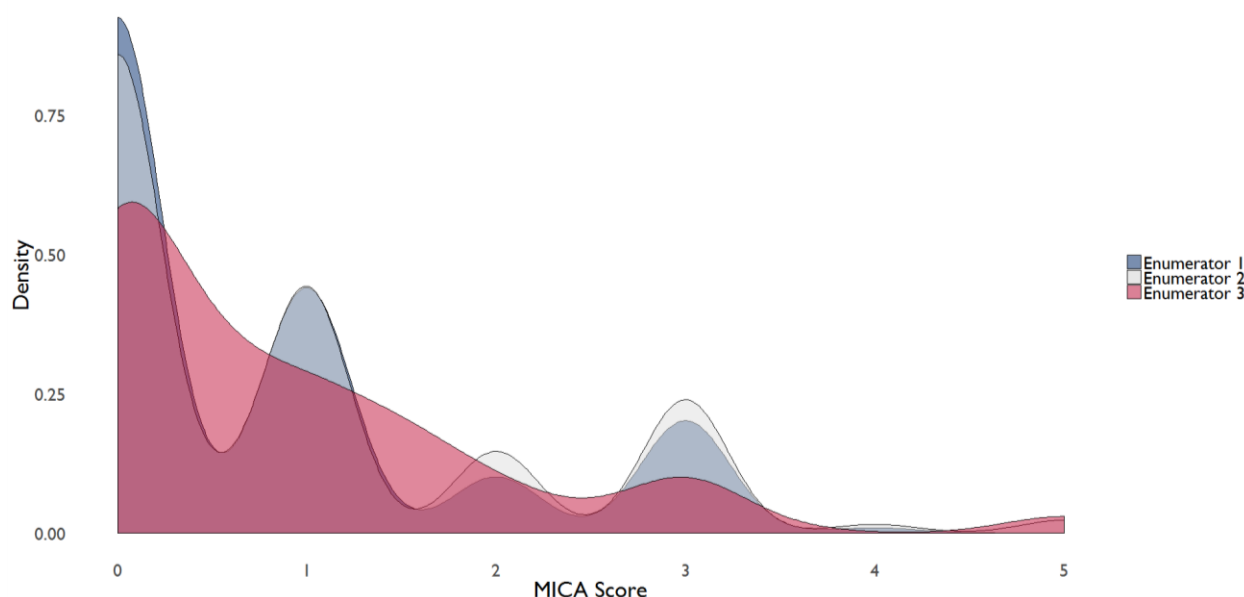
Municipality	Number of Ratings
Ain Albasha	2
Al Seru	2
Alauion	2
Alsharah	2
Alyarmook Aljadedah	2
Bab Amman	3
Iel Jadedah	2
Jizah	2
Manshiat Bani Hasan	3
Mazar Jadedah	2
Mo'ath bin Jabal	2
Shafa	3
Tabaqat Fahl	2
Talal Aljadedah	3

If the evaluation team were perfectly harmonized and had interpreted local conditions in each municipality identically, we would expect all of the MICA scores to be the same. In some cases, the guidance provided was ambiguous or there were inconsistencies in the rubric. For example, for services item 3.1, related to Urban Planning, many municipalities have aspects of a 0 score, as well as a 2 score. Some municipalities may have staff to carry out urban planning and have a plan, but the staff are not actually qualified to do this task. In other cases, there may be an urban plan that is decades old or included a large amount of land designated as “unregulated” with little explanation. Any assessment tool will encounter issues like this given that there will likely always be ambiguity on the ground that does not fit nicely into one single category. The IRR analysis can help systematically flag these issues and then allow for appropriate corrections.

The MICA tool also allows for intermediate scores, e.g. 1.5, to allow for some flexibility in the rating system. As shown in the figure below, there was general agreement between the evaluation team’s MICA enumerators, particularly on the higher end of the MICA scale (e.g. scores of 4 and 5). However, the figure also shows that Enumerator 3 utilized half scores more often than the other two team members.

FIGURE 24: DISTRIBUTION OF MICA SCORES BY ENUMERATOR

There was generally overlap across scores within the evaluation team, but some members utilized half scores more often than others



The lack of overlap suggests that more investigation is needed to analyze the level of reliability. To understand how harmonized scores were, the evaluation team looked at the MICA services scores, i.e. the main MICA items that directly relate to municipality capacity and service provision, MICA GESI scores, the scores that capture how well a municipality has addressed accessibility and gender issues, and the combined GESI-services scores across the full MICA tool. Splitting the data across these three categories allows the evaluation team to highlight gaps in understanding within the team and assess what may be driving divergence in scores. For each of the three data sources – GESI, Services, or the full MICA – the evaluation team performed two tests to assess reliability: percent agreement and intra-cluster correlation.

Percent Agreement

This is a basic statistic showing the percentage of agreement between two or more enumerators. For each municipality, the number of identical scores for each MICA item is counted and divided by the total number of MICA items. For example, in the table below, we see that enumerator 1 gave the municipality a score of two on item 3.4, “Building licensing and control,” while Enumerator 3 entered a score of three, but they both gave a score of zero on item 4.1.

Table 17: Select MICA items with Scoring Divergence

Item	Enumerator 1	Enumerator 3
3.2	1	1
3.3	1	1.5
3.4	2	3
4.1	0	0
4.2	0	0.5
4.3	1	1.5

Percent agreement is a useful first step for assessing how similar ratings for each municipality are overall. One limit of this approach is that it does not account for the fact that raters may, through random chance, have the same scores on some items. If taken as the sole measure of reliability, percent agreement may lead one to conclude there was more synthesis of scores than actually occurred. The evaluation team was cautious about overstating reliability but included percent agreement in its tests because it is an intuitive way to highlight areas of divergence.

Intra-Cluster Correlation

Intra-cluster correlation (ICC) provides an estimate of the relationship between scores within some grouping, such as municipalities. The ICC provides an estimate of how much variability in scores is a result of the variance between raters. That is, ICC provides a measure of the similarity of scores for the same municipality after taking into account the variation across all of the MICA items. A high ICC coefficient denotes stronger similarity between raters, while a lower coefficient suggests a weaker relationship. This measure is useful for understanding how similar raters were to each other while taking into account the fact that there is going to be variability across items since some categories are more likely to be within the control of municipalities and garner a higher or lower score than other MICA items.

The appropriateness of the ICC depends on whether one considers the MICA scoring options to be categorical or continuous. Although the ODK MICA tool included the option for scores that were between integer item values, e.g. 1.5, and the CITIES team provided guidance that enumerators should score outside of the integer categories for municipalities on the margin of one score or another, the tool is set up with clear categorical item scores. For example, not all MICA items run from zero to five, some MICA items go from 0 to 1 to 5 and only have associated guidance for those integer items. Regardless, the MICA scores are numeric and ICC coefficients were calculated based on the level of agreement.

FULL MICA SCORE RELIABILITY

The evaluation team assessed the overall inter-rater reliability across MICA scores. The MICA tool is constructed to allow for more ambiguity on service-related scores and less on GESI-related scores. By seeing how harmonized enumerator scoring was across all items, the evaluation team was able to understand where there may have been general misunderstanding about certain municipality conditions. This broader reliability analysis was the first step in a more detailed investigation of enumerator harmonization.

As shown in the table below, when both GESI and Services scores are included together, the agreement statistics tend to be higher. One reason for this is the low variance for GESI scoring—the evaluation team raters were generally synthesized in their scoring for these items as shown in the high percent agreement in Table 16. The ICC levels are also high given this low variation between and across items and raters when GESI scores are included. The high ICCs are driven by the variance across clusters. For example, Alsharah's raters were not identical, but the magnitude of their scoring was very similar, resulting in an ICC of 0.99.

TABLE 18: OVERALL MICA SCORE RELIABILITY STATISTICS

Municipality	Percent Agreement	ICC
Alsharah	86.05	0.99
Al Seru	83.72	0.87
Ain Albasha	79.07	0.96
Alauion	76.74	0.85
Bab Amman	76.74	0.97
Talal Aljadedah	76.74	0.99
Mo'ath bin Jabal	71.43	0.94
Shafa	69.77	0.99
Alyarmook Aljadedah	69.23	0.94
Tabaqat Fahl	69.23	0.98
Iel Jadedda	67.44	0.93
Manshiat Bani Hasan	66.67	0.97
Mazar Jadedda	65.12	0.99
Jizah	62.79	0.89

SERVICE SCORE RELIABILITY

The MICA Service items cover 23 of the municipal responsibilities outlined in Municipalities Law 41, as well as other relevant services such as public utilities, transportation, or disaster management where the municipality plays some coordinating role in conjunction with the relevant ministry or public authority. For each service, quality levels are assigned scores from zero to five. The evaluation team met with CITIES as part of the MICA training to discuss these items as well as during the evaluability assessment and design stages to understand the service aspect of the MICA tool. The evaluation team also had external subject matter experts review the MICA and provide additional input and research to help the team further understand the service-related MICA items.

TABLE 19: MICA SERVICE SCORE RELIABILITY STATISTICS

Municipality	Percent Agreement	ICC
Al Seru	73.91	0.84
Alsharah	73.91	0.83
Ain Albasha	60.87	0.88
Alauion	60.87	0.69
Talal Aljadedah	60.87	0.75
Alyarmook Aljadedah	56.52	0.87
Bab Amman	56.52	0.84
Manshiat Bani Hasan	56.52	0.73
Shafa	52.17	0.78
Mo'ath bin Jabal	47.83	0.7
Tabaqat Fahl	47.62	0.64
Iel Jadedda	43.48	0.75
Jizah	43.48	0.30
Mazar Jadedda	43.48	0.80

As shown in the table above, there was limited initial agreement between enumerators on the MICA services scores. The average percent agreement was about 56 percent. Taken together, the reliability statistics above suggest that the enumerators were consistently in disagreement on the appropriate services scores. The services reliability tests prompted the evaluation team to revisit divergent scores and determine a final score for each municipality. This process involved a thorough review of the enumerator notes and discussion of each item to see if there may have been details that an enumerator missed. As noted in the evaluation design proposal, there are certain limits to the MICA tool. The MICA scoring

process is not only sensitive to enumerator bias, but also to the availability of local officials and their willingness to share information. Some variation in scores is expected, but the results of the evaluation team's initial analysis prompted additional follow-up as discussed below.

GESI SCORE RELIABILITY

The 17 GESI items are scored on a binary, yes-no scale. The guidance on the GESI items is generally clear, so more similarity in scores was expected. Given the binary scale, ICC is not generally recommended as the most appropriate measure of inter-rater reliability. The inclusion of ICC can still provide an edifying comparison for understanding how this statistic varies in relation to percent agreement and how related scores are between and across clusters, or municipalities. The GESI scores were on average about 93 percent in agreement, suggesting strong reliability. Jizah had a relatively low percent agreement, at about 84 percent, which prompted more investigation into where enumerators may have seen different GESI-related municipality infrastructure or interpreted the availability of some service differently. The high ICC presented in the table below is largely a function of the binary nature of GESI scores; that is, scores will be inherently related to each other based on the set-up of the measure.

TABLE 20: MICA GESI SCORE RELIABILITY STATISTICS

Municipality	Percent Agreement	ICC
Ain Albasha	100.00	1.00
Alsharah	100.00	1.00
Bab Amman	100.00	1.00
Mo'ath bin Jabal	100.00	1.00
Al Seru	94.74	0.89
Alauion	94.74	0.95
Iel Jadedah	94.74	0.99
Talal Aljadedah	94.74	1.00
Tabaqat Fahl	94.12	1.00
Ma'an	89.47	0.99
Mazar Jadedah	89.47	0.99
Shafa	89.47	1.00
Alyarmook Aljadedah	86.67	0.96
Jizah	84.21	0.98
Manshiet bani Hani	77.77	0.98

SYNTHESIZING IE MICA SCORES

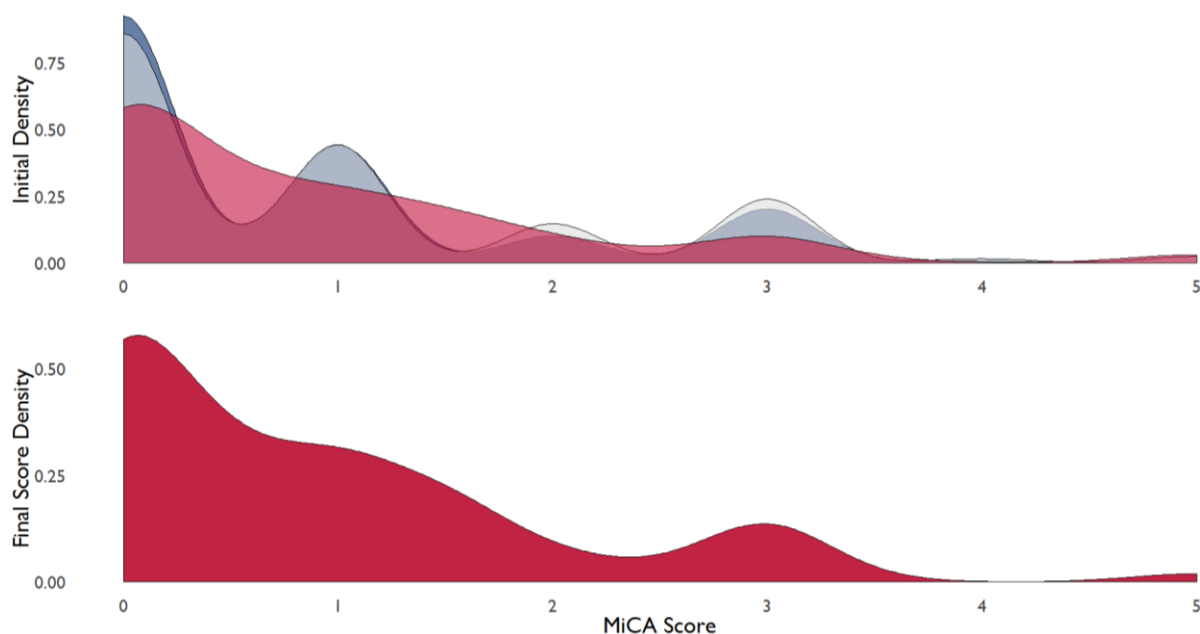
Based on the above analysis, and as part of overall quality assurance, the evaluation team undertook a MICA synthesis process to align scores across raters. The end goal of this was to develop a final dataset of comparison municipality MICA scores. The evaluation team went through each item-by-item disagreement for each municipality, compared scores, notes, and then reviewed the MICA guidance to develop a final score. As previously mentioned, in some cases, the guidance was ambiguous or inconsistent. In other cases, the enumerators simply went over their notes, discussed what they observed, and determined whether to maintain or change the final score. The table below shows the same items for the same municipality from the start of this IRR section. The last column shows the resolved scores. For services item 3.3, which covers the participation of the citizens and the business community in identifying citizens' needs, the enumerators resolved that the frequency of such activities did not warrant a score above a 1, denoting occasional activities.

TABLE 21: SELECT MICA ITEMS WITH SCORING DIVERGENCE AND FINAL SCORING

Item	Enumerator 1	Enumerator 3	Final
3.2	1	1.0	1.0
3.3	1	1.5	1.0
3.4	2	3.0	3.0
4.1	0	0.0	0.0
4.2	0	0.5	0.5
4.3	1	1.5	1.5

Each municipality had an average of 3.8 points change across all MICA services items. Jizah municipality saw the most changes between the lead enumerator and the final score, with a total of seven score updates. Despite around four points changing in each municipality, the average change in the lead enumerator's score was only 0.03 points. In eight cases, scores changed by one point, scores changed by 1.5 points in three cases, and in two cases the MICA services scores changed by 2.5 points. The two largest changes in the final scores were in Shafa and Jizah. In Shafa, 2.5 points were removed for item 4.2, which covers maintenance of assets, while in Jizah 2.5 points were added to item 4.7, which covers public transportation. The Jizah score was changed as the team learned that the municipality is connected to a bus line, but that this line does not consistently run within the municipality. For Shafa, the team noted that while there is local maintenance, it is almost entirely outsourced and not provided through municipality. Despite these changes, the final MICA scores generally preserved the lead enumerator's initial assessment. The figure below shows the initial distribution of scores and the final MICA score set, which has the same general distribution as the lead enumerator's scores.

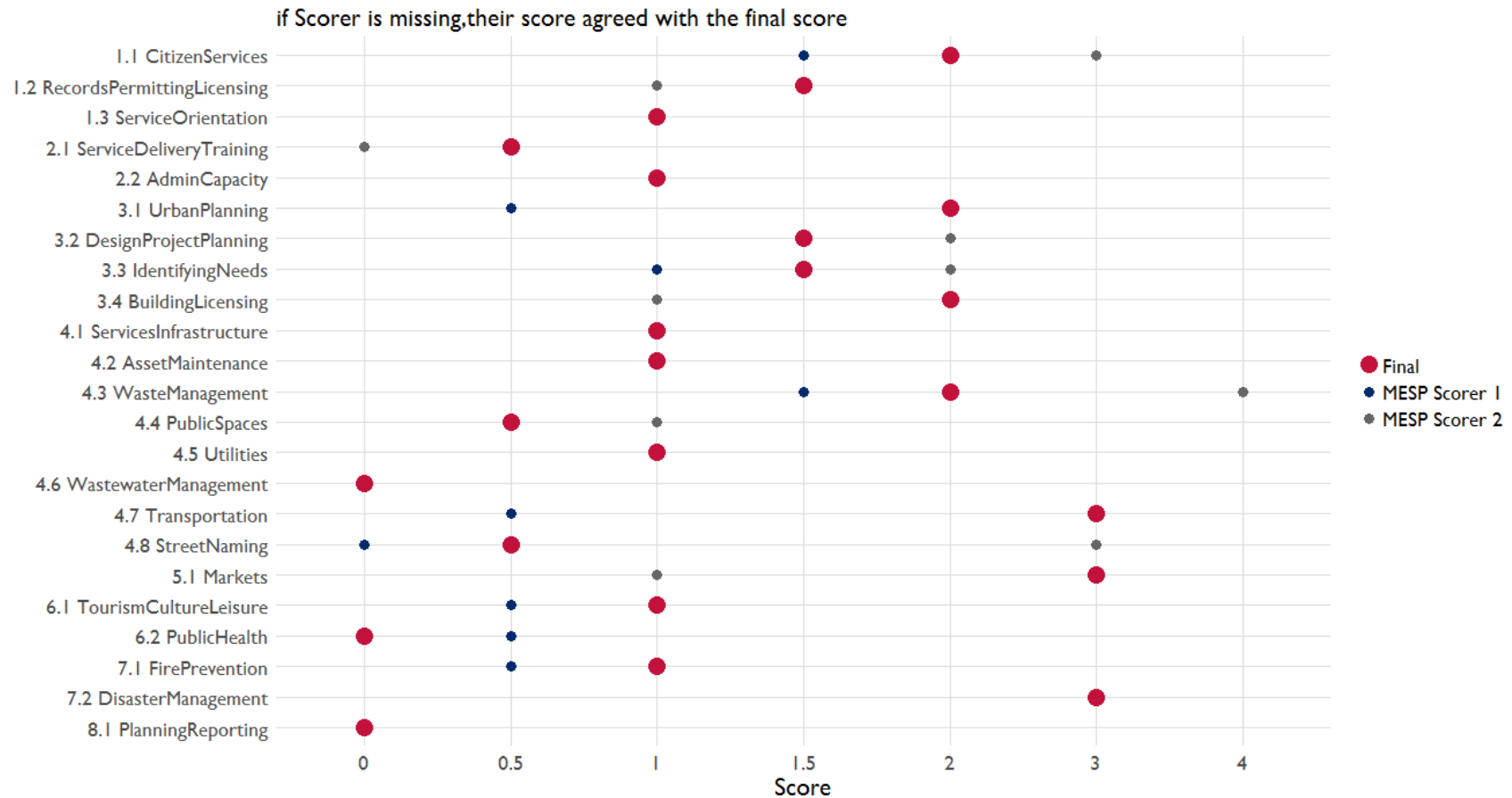
FIGURE 25: DISTRIBUTION OF INITIAL AND FINAL MICA SCORES



At the municipality level, the evaluation team generated the numeric error between raters and ranked each municipality to see how internally consistent the scoring was across enumerators. The evaluation team found that the average error across the 15 municipalities was around 0.3 points.

The graph below shows the example of Jizah, which had seven items for which there was disagreement between enumerators. Although this municipality had one of the largest point disagreements for a single MICA item, the graph below highlights the fact that, on many of the items, the disagreement was the difference of 0.5 points, and there was no disagreement in nine cases.

FIGURE 26: CHANGE IN MICA SERVICES SCORES FOR JIZAH



Results for the GESI items were similar, although with fewer changes overall given the low number of GESI observations. There were seven changes to GESI scores across the 15 comparison municipalities. The reason for these changes was similar to that of the services scores: the tool itself allows for subjective judgment. This is particularly true for the binary GESI items. The figure below illustrates the limited number of GESI scores and subsequent changes.

FIGURE 27: CHANGES IN MICA GESI SCORES FOR MAZAR JADEDA



From this review, the evaluation team was able to create a final dataset of MICA scores to use for descriptive analysis and to test balance between assignment groups.

ANNEX 3: INTER RATER AGREEMENT, BY MUNICIPALITY

FIGURE 28 AGREEMENT ERROR BY ITEM - AIN ALBASHA

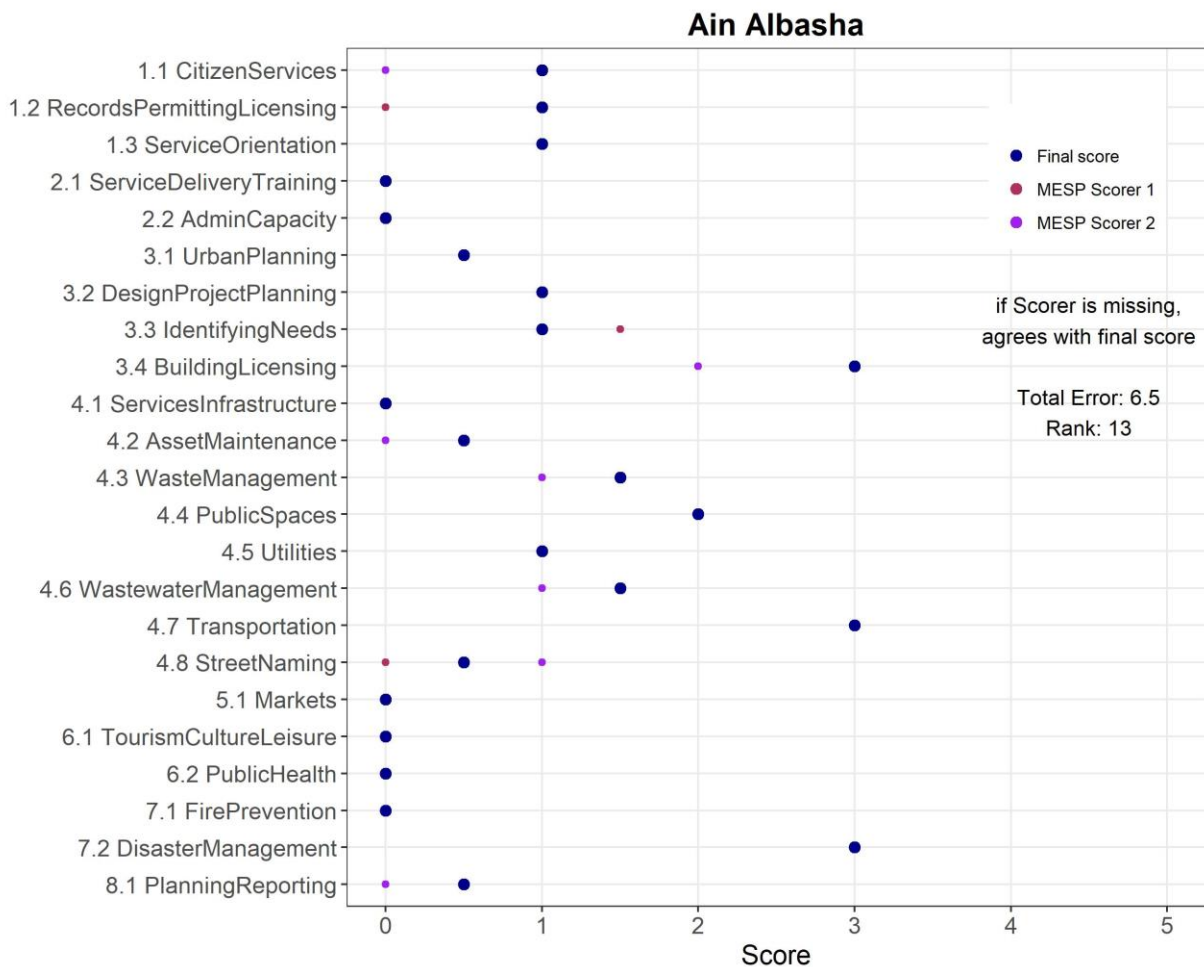


FIGURE 29 AGREEMENT ERROR BY ITEM - AL SERU

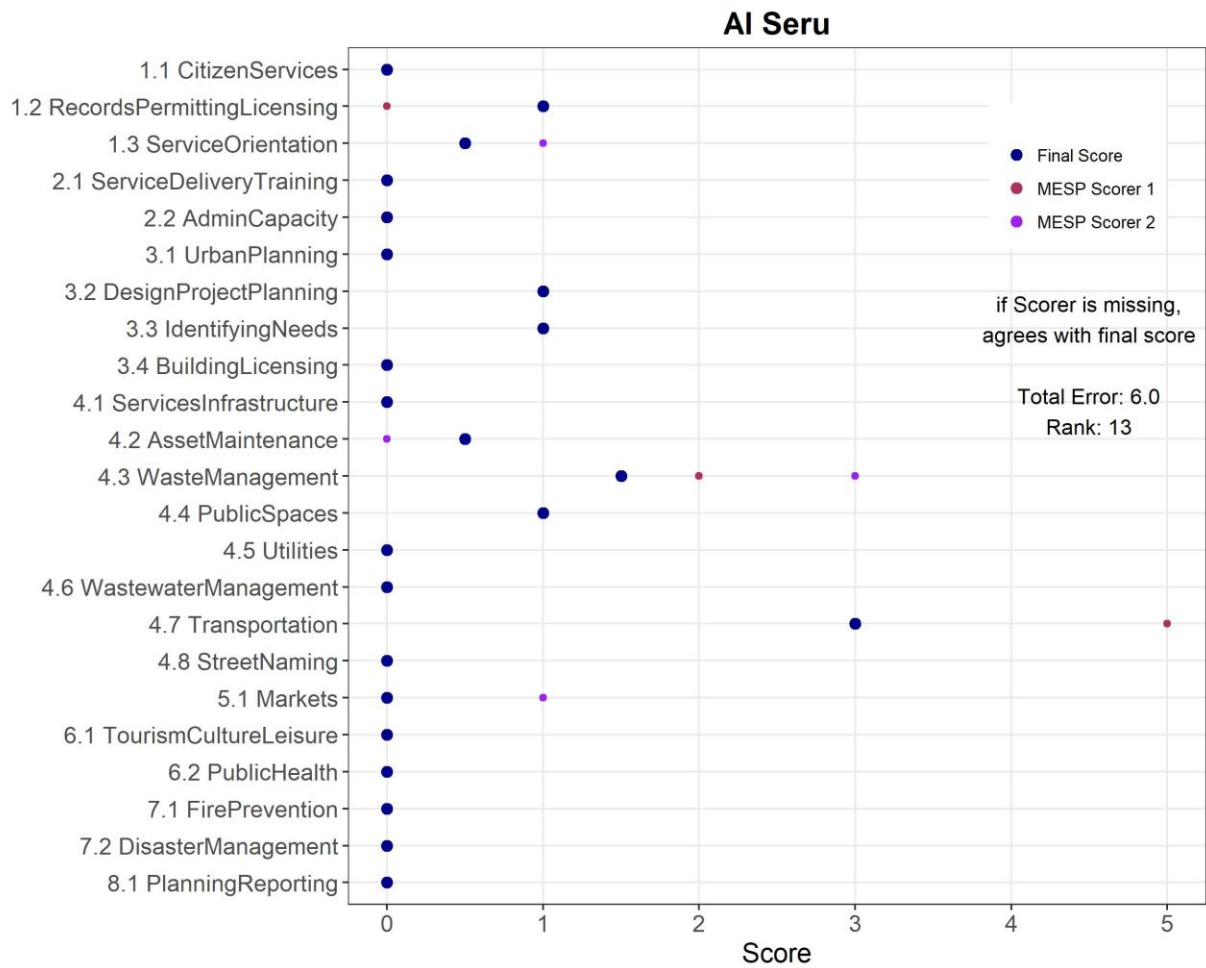


FIGURE 30 AGREEMENT ERROR BY ITEM – ALAUION

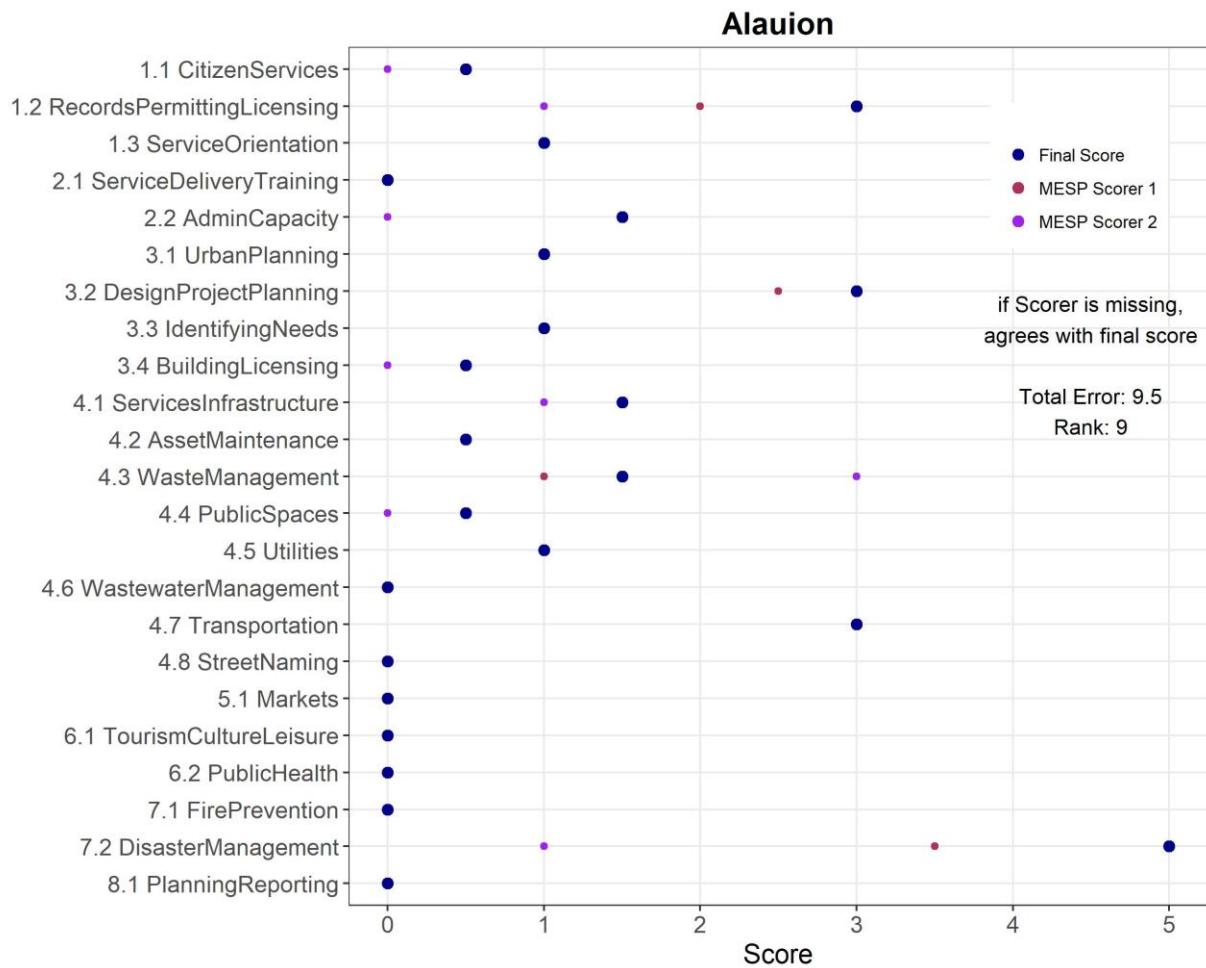


FIGURE 3I AGREEMENT ERROR BY ITEM – ALSHARAH

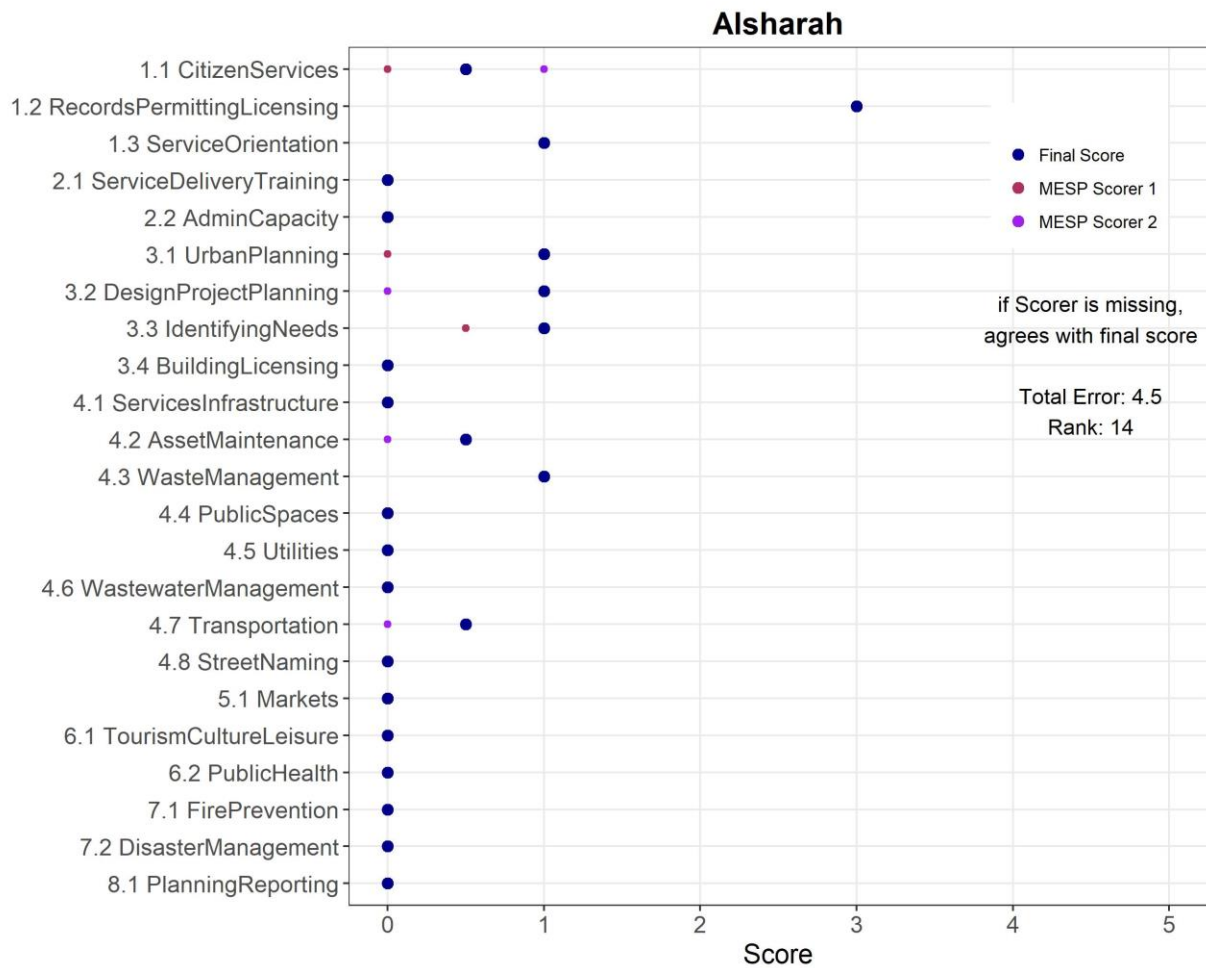


FIGURE 32 AGREEMENT ERROR BY ITEM - ALYARMOOK ALJADEDAH

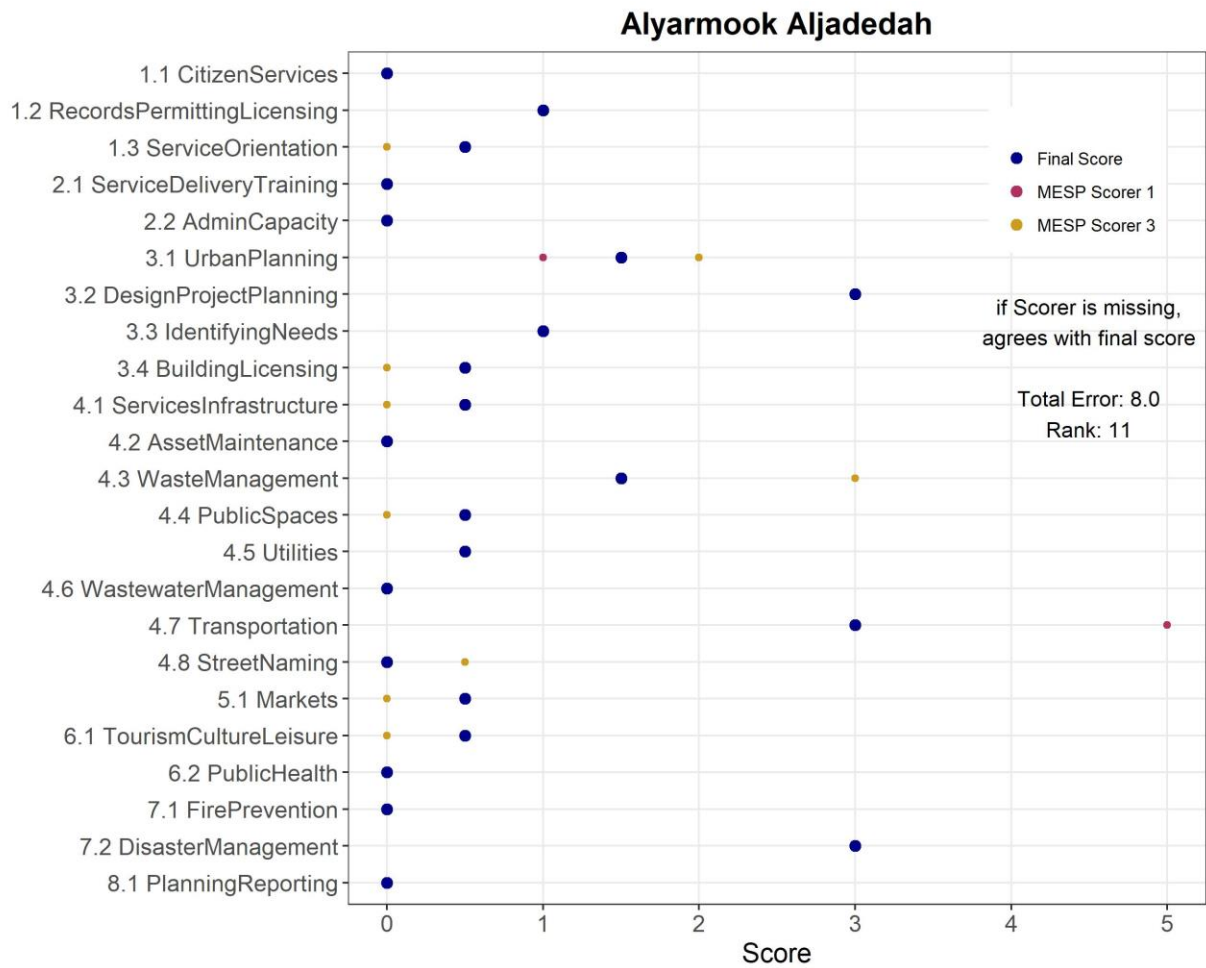


FIGURE 33 AGREEMENT ERROR BY ITEM - BAB AMMAN

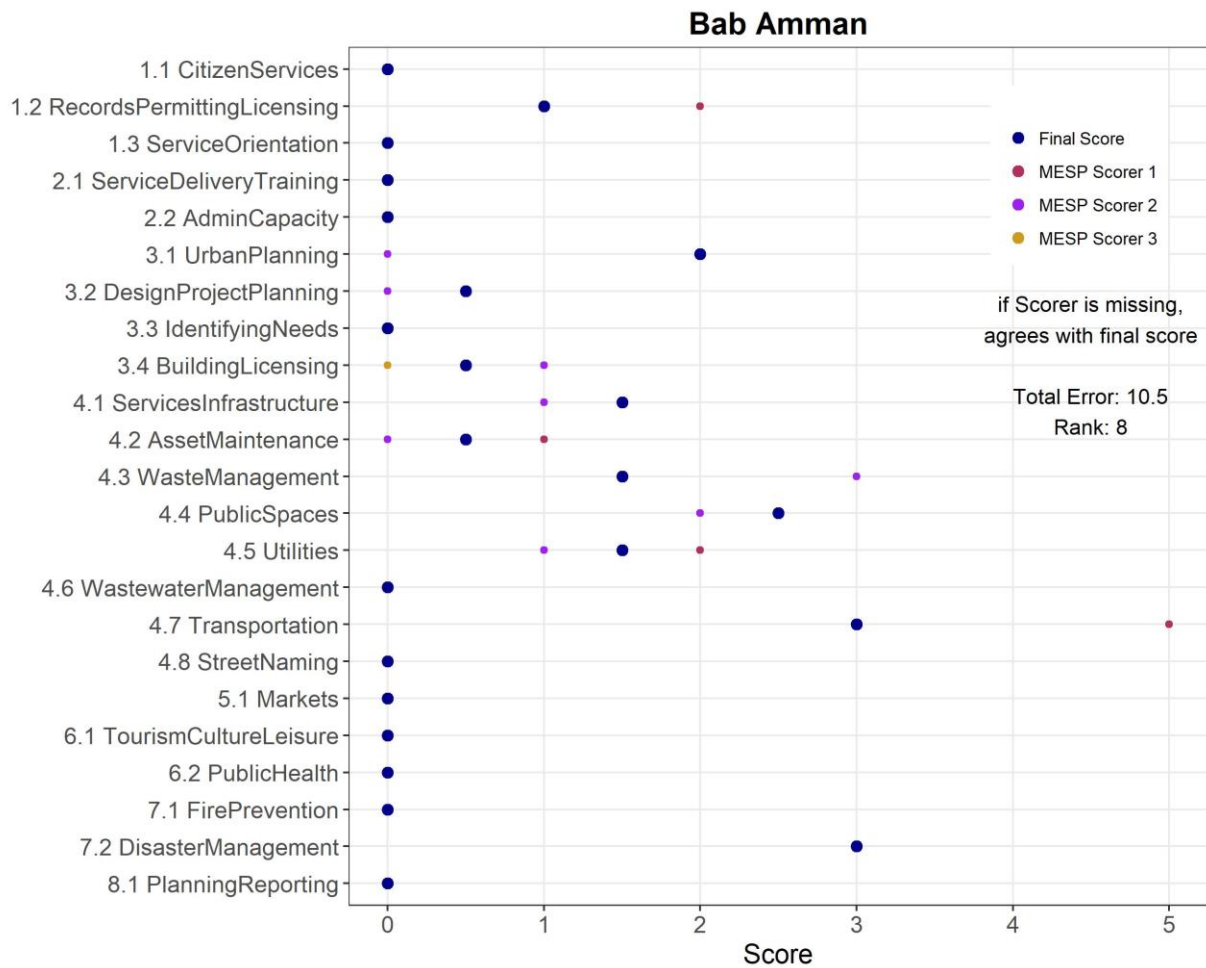


FIGURE 34 AGREEMENT ERROR BY ITEM - IEL JADEDAAH

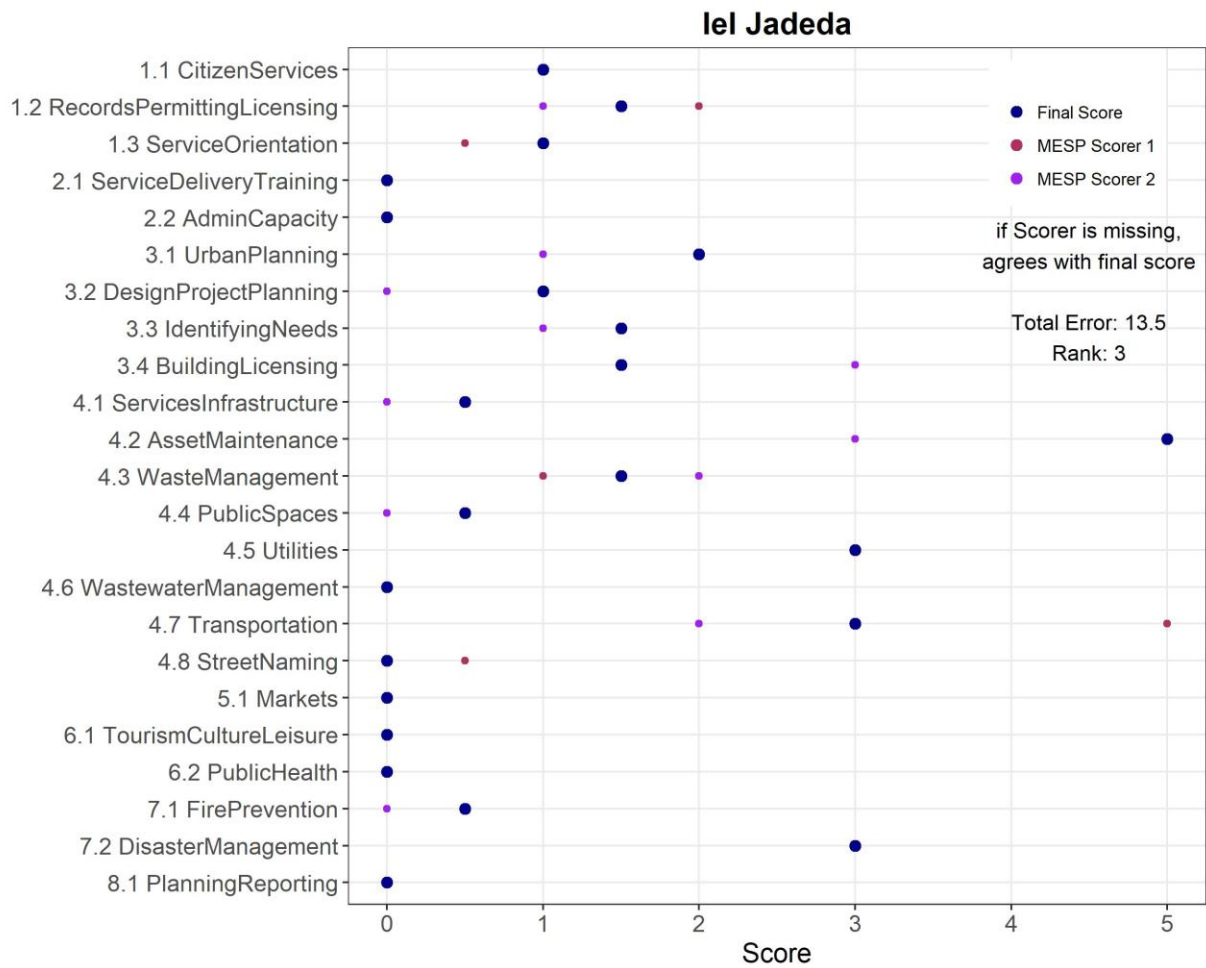


FIGURE 35 AGREEMENT ERROR BY ITEM - JIZAH

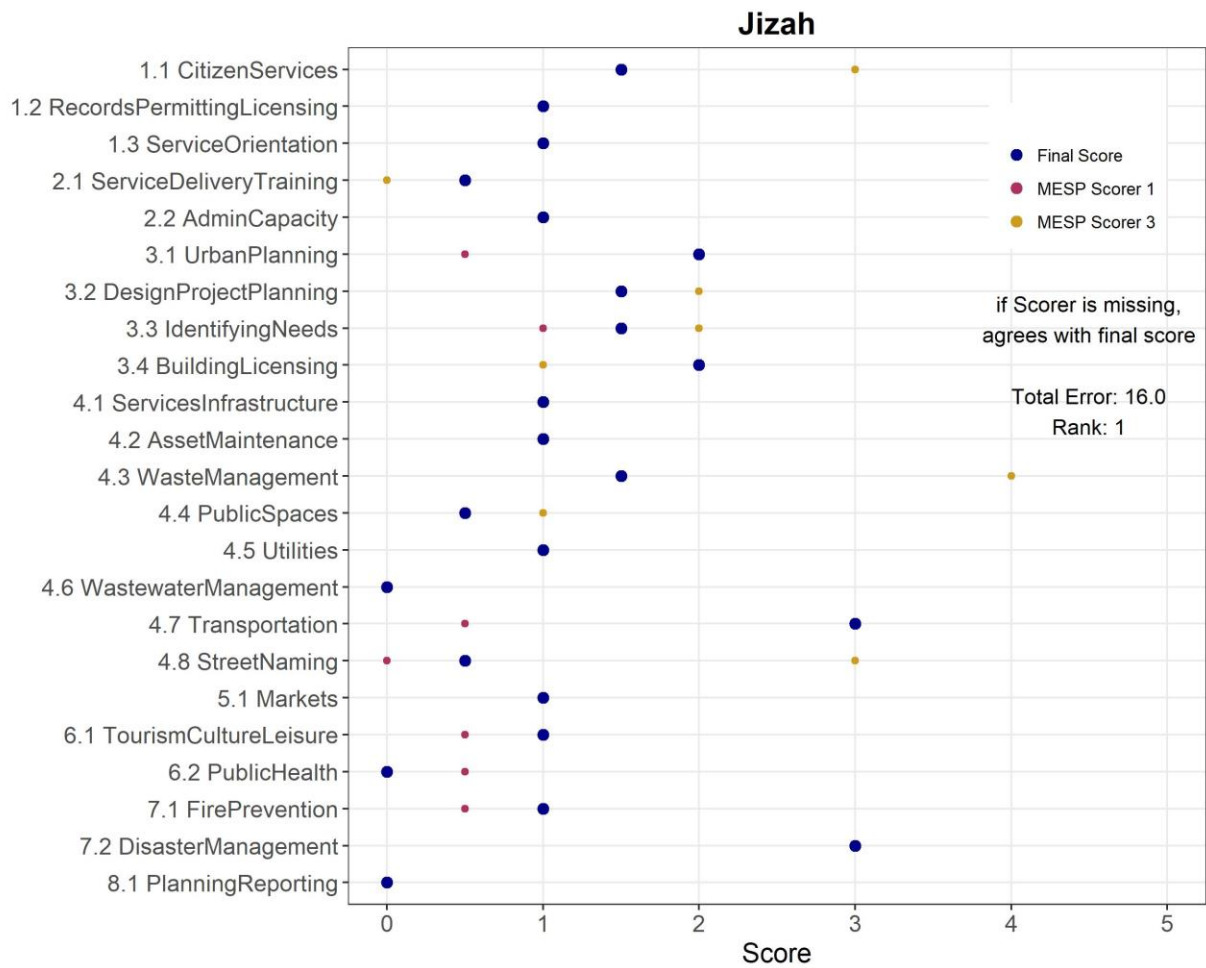


FIGURE 36 AGREEMENT ERROR BY ITEM - MANSHIAT BANE

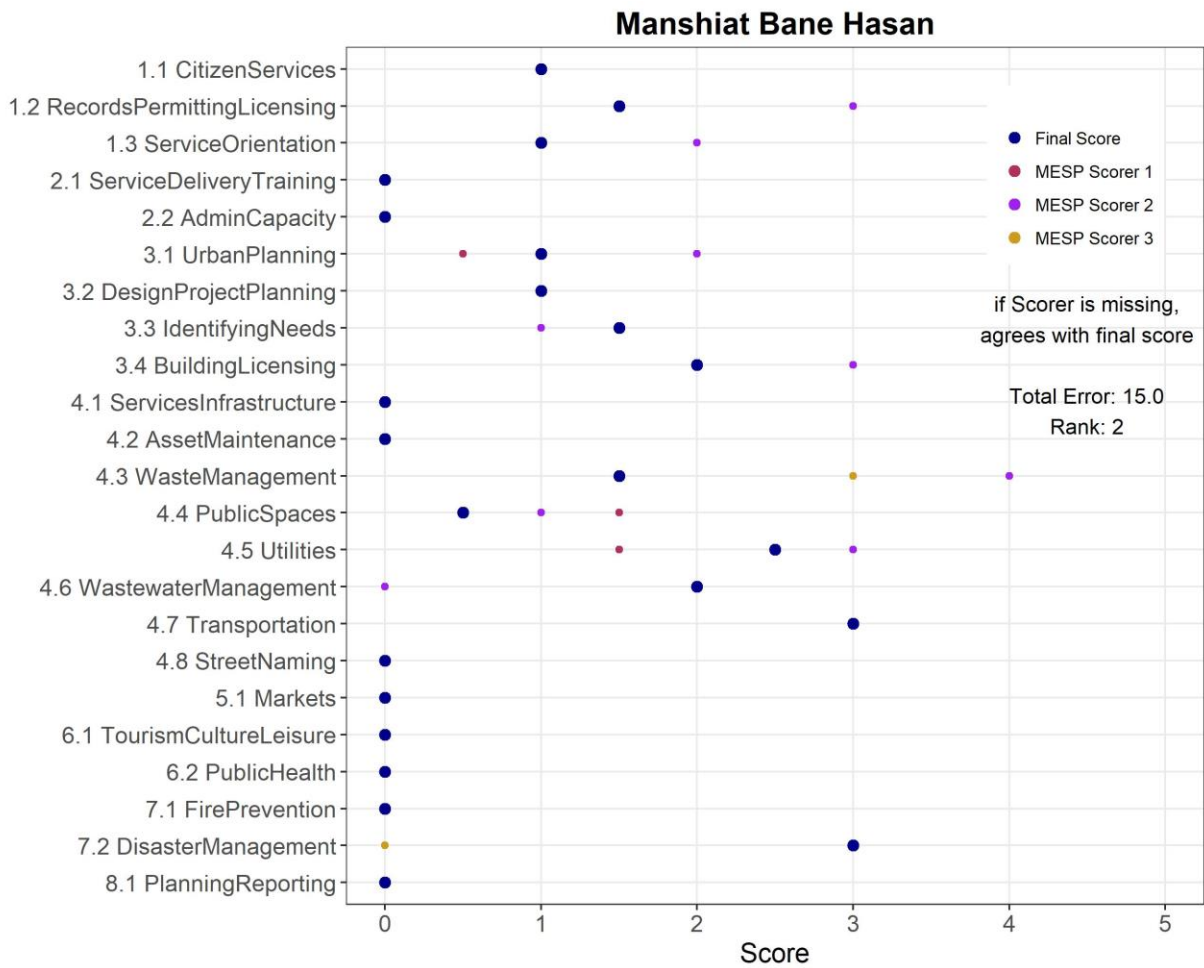


FIGURE 37 AGREEMENT ERROR BY ITEM - MAZAR JADEDA

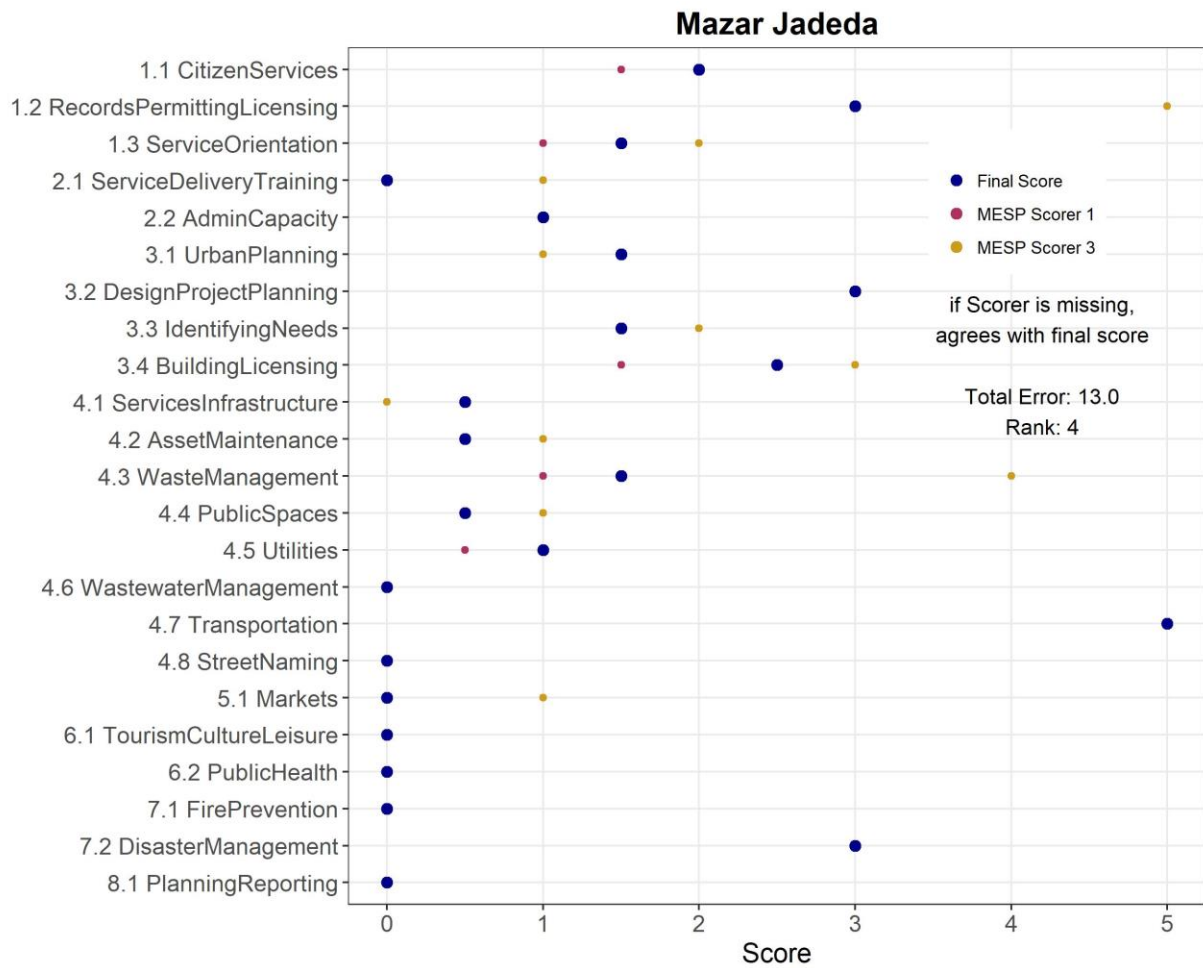


FIGURE 38 AGREEMENT ERROR BY ITEM - MO'ATH ABN JABAL

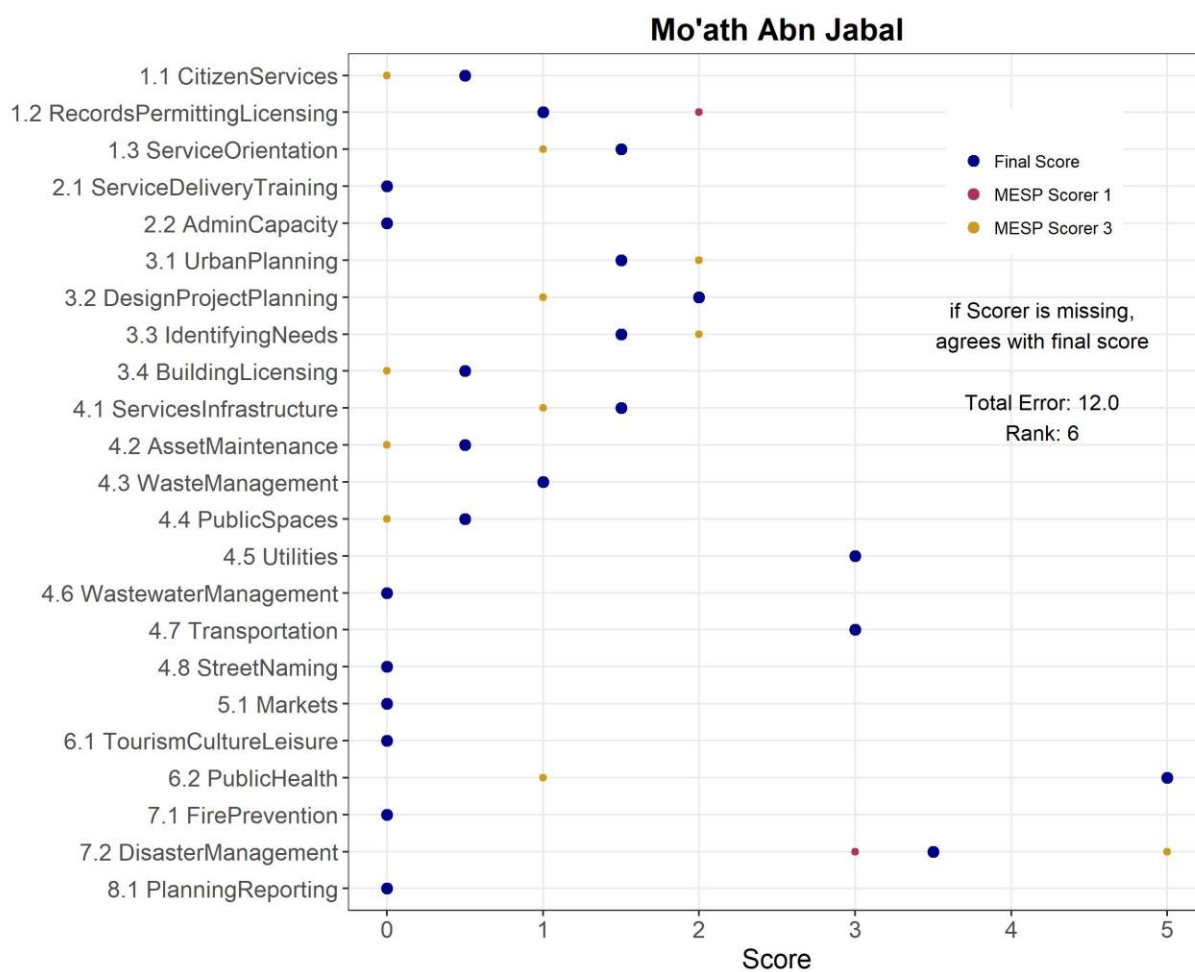


FIGURE 39 AGREEMENT ERROR BY ITEM – SHAFa

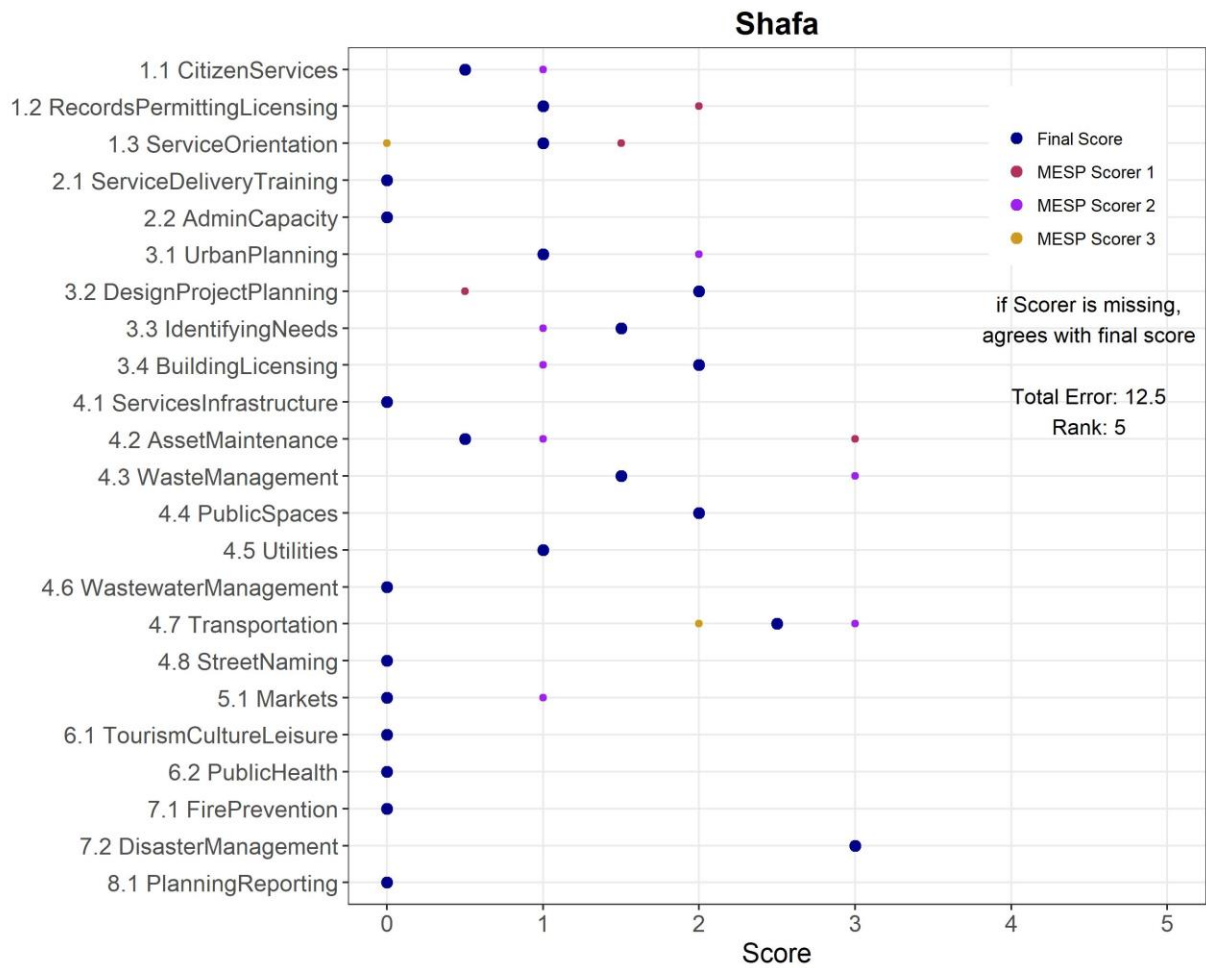


FIGURE 40 AGREEMENT ERROR BY ITEM - TABAQAT FAHL

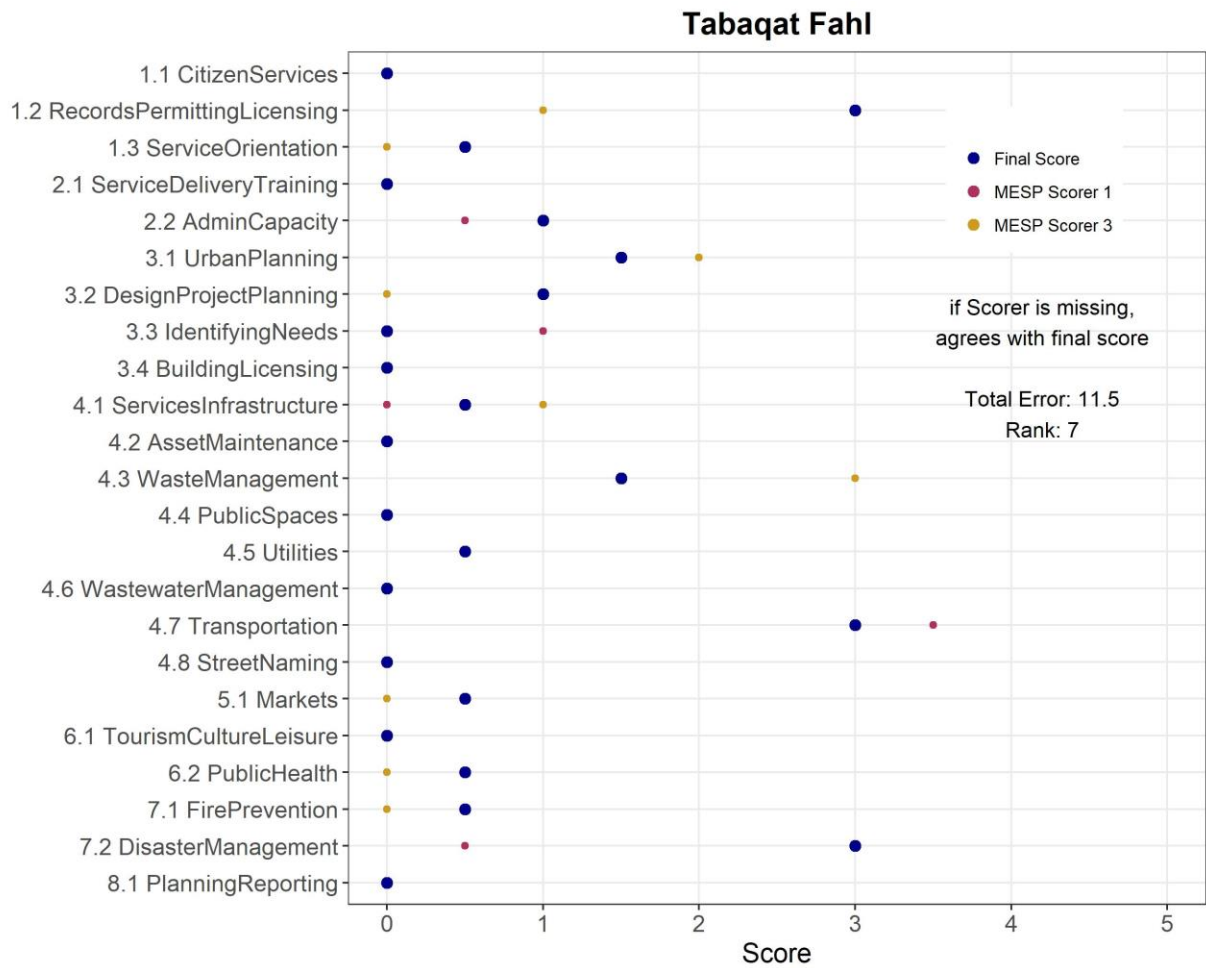
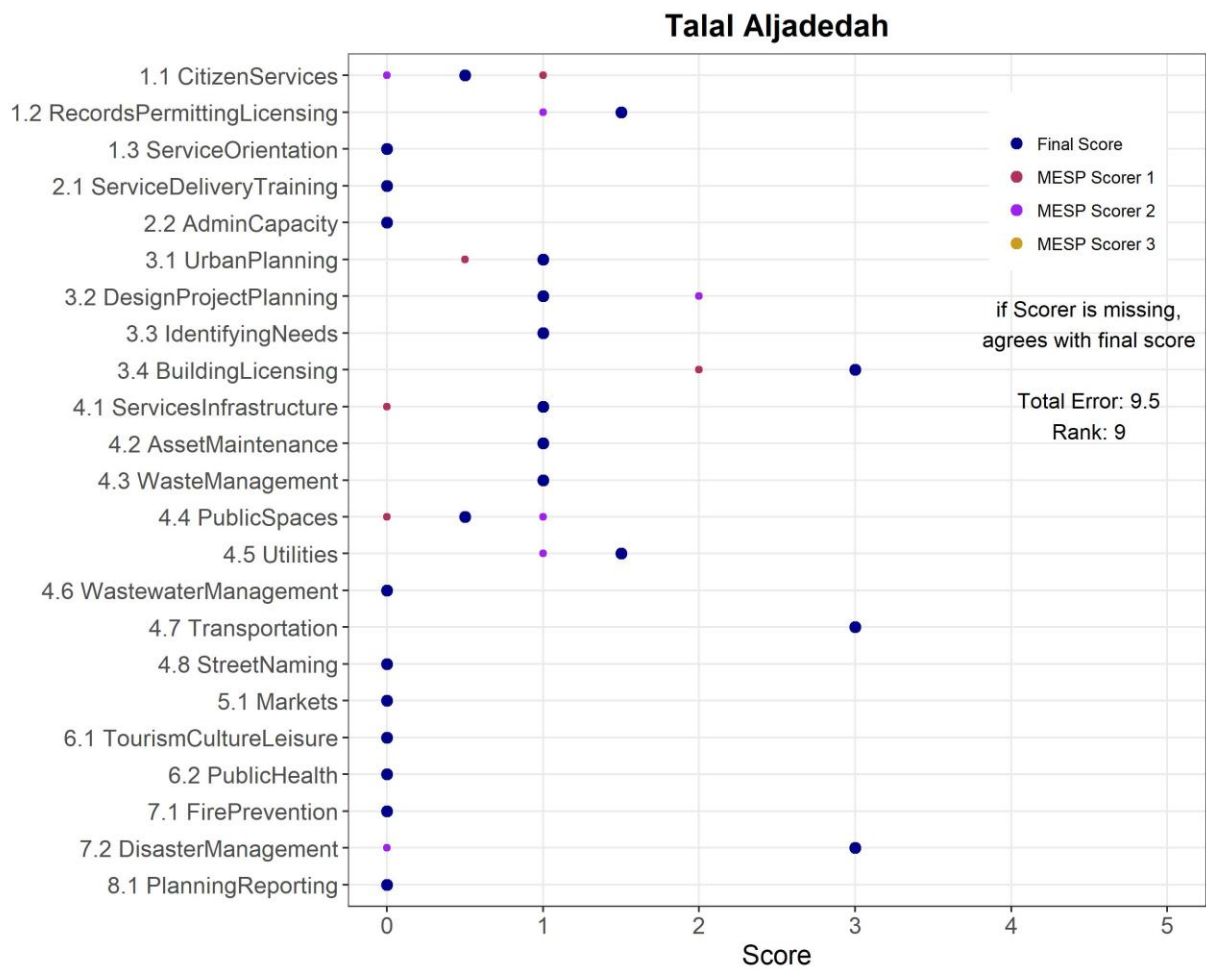


FIGURE 4I AGREEMENT ERROR BY ITEM - TALAL ALJADEDAH



ANNEX 4: INVESTIGATING BASELINE EQUIVALENCE

The tables below show the full MICA services and GESI items and the balance statistics for each of the following configurations:

- **Full configuration:** All 33 CITIES municipalities and 15 evaluation municipalities
- **Partial Configuration:** 23 Class B and C CITIES municipalities and the evaluation municipalities
- **Matched Configuration:** 15 matched CITIES municipalities and the evaluation municipalities.

As mentioned in the body of the report, mean differences at or below an absolute value of 0.25 generally suggest strong balance. The upper and lower confidence intervals used the normalized mean value and a 95 percent confidence level.

TABLE 22: DETAILED MICA SERVICES ITEMS BALANCE STATISTICS, FULL CONFIGURATION (ALTERNATE MICA SCORES)

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
1.1	1.45	0.80	0.65	-0.66	
1.2	1.14	1.70	-0.56	0.65	
1.3	1.11	0.80	0.31	-0.50	
2.1	0.39	0.10	0.29	-0.59	
2.2	0.68	0.33	0.35	-0.63	
3.1	1.42	1.27	0.16	-0.21	
3.2	1.67	1.77	-0.10	0.10	
3.3	1.27	1.17	0.11	-0.16	
3.4	1.70	1.40	0.30	-0.25	
4.1	0.95	0.60	0.35	-0.50	
4.2	1.36	0.77	0.60	-0.50	
4.3	2.11	1.80	0.31	-0.34	
4.4	1.58	0.80	0.78	-0.89	
4.5	1.83	1.23	0.60	-0.56	
4.6	0.59	0.23	0.36	-0.49	
4.7	2.58	2.93	-0.36	0.33	
4.8	0.58	0.07	0.51	-0.83	
5.1	1.91	2.40	-0.49	0.33	

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
6.1	0.79	0.13	0.65	-0.71	
6.2	1.29	0.50	0.79	-0.54	
7.1	0.06	0.20	-0.14	0.44	
7.2	0.70	1.57	-0.87	0.86	
8.1	0.47	0.03	0.44	-0.70	

TABLE 23: DETAILED MICA SERVICES ITEMS BALANCE STATISTICS, PARTIAL CONFIGURATION (ALTERNATE MICA SCORES)

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
1.1	1.22	0.8	0.42	-0.45	
1.2	0.85	1.7	-0.85	1.20	
1.3	0.85	0.8	0.05	-0.10	
2.1	0.13	0.1	0.03	-0.10	
2.2	0.46	0.33	0.12	-0.24	
3.1	1.09	1.27	-0.18	0.31	
3.2	1.52	1.77	-0.24	0.25	
3.3	1.17	1.17	0.01	-0.01	
3.4	1.46	1.4	0.06	-0.05	
4.1	0.72	0.6	0.12	-0.20	
4.2	0.89	0.77	0.12	-0.11	
4.3	1.54	1.8	-0.26	0.41	
4.4	1.24	0.8	0.44	-0.58	
4.5	1.46	1.23	0.22	-0.21	
4.6	0.13	0.23	-0.1	0.19	
4.7	2.13	2.93	-0.8	0.77	
4.8	0.35	0.07	0.28	-0.59	

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
5.1	1.3	2.4	-1.1	0.84	
6.1	0.48	0.13	0.34	-0.44	
6.2	0.98	0.5	0.48	-0.34	
7.1	0.04	0.2	-0.16	0.51	
7.2	0.52	1.57	-1.04	1.10	
8.1	0.59	0.03	0.55	-0.83	

TABLE 24: DETAILED MICA SERVICES ITEMS BALANCE STATISTICS, MATCHED CONFIGURATION (ALTERNATE MICA SCORES)

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
1.1	1.23	0.8	0.43	-0.45	
1.2	0.83	1.7	-0.87	1.24	
1.3	0.93	0.8	0.13	-0.29	
2.1	0.13	0.1	0.03	-0.10	
2.2	0.43	0.33	0.1	-0.2	
3.1	1.03	1.27	-0.23	0.38	
3.2	1.53	1.77	-0.23	0.22	
3.3	1.07	1.17	-0.1	0.15	
3.4	1.63	1.4	0.23	-0.22	
4.1	0.73	0.6	0.13	-0.22	
4.2	0.93	0.77	0.17	-0.14	
4.3	1.5	1.8	-0.3	0.43	
4.4	1.37	0.8	0.57	-0.75	
4.5	1.47	1.23	0.23	-0.23	
4.6	0.13	0.23	-0.1	0.17	

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
4.7	2.33	2.93	-0.6	0.63	
4.8	0.37	0.07	0.3	-0.52	
5.1	1.6	2.4	-0.8	0.59	
6.1	0.6	0.13	0.47	-0.48	
6.2	1.3	0.5	0.8	-0.53	
7.1	0	0.2	-0.2	0.77	
7.2	0.67	1.57	-0.9	0.93	
8.1	0.73	0.03	0.7	-0.89	

TABLE 26: DETAILED MICA GESI ITEMS BALANCE STATISTICS, PARTIAL CONFIGURATION (ALTERNATE MICA SCORES)

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
1.1	0.09	0.00	0.09	1.73	
1.2	0.13	0.20	-0.07	0.50	
1.3	0.04	0.00	0.04	1.65	
2.1	0.09	0.00	0.09	1.73	
2.2	0.00	0.00	0.00	0.00	
3.1	0.35	0.13	0.21	1.11	
3.2	0.30	0.20	0.10	0.52	
3.3	0.00	0.00	0.00	0.00	
3.4	0.17	0.00	0.17	1.91	
4.1	0.00	0.00	0.00	0.00	
4.2	0.00	0.00	0.00	0.00	
4.3	0.22	0.00	0.22	2.02	
4.4	0.00	0.20	-0.20	0.84	

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
4.5	0.09	0.00	0.09	1.73	
4.6	0.00	0.00	0.00	0.00	
4.7	0.00	0.00	0.00	0.00	
4.8	0.04	0.00	0.04	1.65	
5.1	0.09	0.00	0.09	1.73	
6.1	0.13	0.20	-0.07	0.50	
6.2	0.04	0.00	0.04	1.65	
7.1	0.09	0.00	0.09	1.73	
7.2	0.00	0.00	0.00	0.00	
8.1	0.35	0.13	0.21	1.11	

TABLE 27: DETAILED MICA GESI ITEMS BALANCE STATISTICS, MATCHED CONFIGURATION (ALTERNATE MICA SCORES)

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
1.1	0.07	0.00	0.07	2.00	
1.2	0.07	0.20	-0.13	1.12	
1.3	0.00	0.00	0.00	0.00	
2.1	0.13	0.00	0.13	2.15	
2.2	0.00	0.00	0.00	0.00	
3.1	0.40	0.13	0.27	1.40	
3.2	0.27	0.20	0.07	0.35	
3.3	0.00	0.00	0.00	0.00	
3.4	0.20	0.00	0.20	2.33	
4.1	0.00	0.00	0.00	0.00	
4.2	0.00	0.00	0.00	0.00	

Measure	Treatment Mean	Comparison Mean	Difference	Standardized Difference	Mean
4.3	0.27	0.00	0.27	2.55	
4.4	0.00	0.20	-0.20	0.67	
4.5	0.07	0.00	0.07	2.00	
4.6	0.00	0.00	0.00	0.00	
4.7	0.00	0.00	0.00	0.00	
4.8	0.07	0.00	0.07	2.00	
5.1	0.07	0.00	0.07	2.00	
6.1	0.07	0.20	-0.13	1.12	
6.2	0.00	0.00	0.00	0.00	
7.1	0.13	0.00	0.13	2.15	
7.2	0.00	0.00	0.00	0.00	
8.1	0.40	0.13	0.27	1.40	

ANNEX 5: GENERAL POPULATION SURVEY

The following table breaks down sample size by municipality and CITIES treatment status.

Table 28: Household survey sample size by municipality, CITIES treatment status

Region	Governorate	Municipality	Class	Treatment	Sample
Central	Amman	Amman	A	0	969
Central	Amman	Jizah	B	0	173
Central	Amman	Na'oor	B	0	257
Central	Amman	Um Al-Rasas	C	0	70
Central	Balqa	Ain Albasha	B	0	435
Central	Balqa	Alshoneh Alwasta	B	0	205
Central	Balqa	Fuhais	B	0	115
Central	Balqa	Mahes	C	0	118
Central	Balqa	Swaimeh	D	0	78
Central	Madaba	Madaba Alkubrah	A	1	462
Central	Madaba	Jabal bne Hamedah	C	0	130
Central	Zarqa	Zarqa	A	1	982
Central	Zarqa	Al Hashemiyah	B	0	283
Central	Zarqa	Bierain	C	0	126
Central	Zarqa	El-Hallabat	C	1	117
North	Ajlun	Ajlun Alkubrah	A	1	301
North	Ajlun	Janed	B	0	168
North	Ajlun	Shafa	B	0	131
North	Irbid	Irbid Alkubrah	A	1	970
North	Irbid	Dair Abi Sa'id Jadedda	B	0	121
North	Irbid	Mazar Jadedda	B	0	119
North	Irbid	Rabeat Al Koorah	B	0	125

Region	Governorate	Municipality	Class	Treatment	Sample
North	Irbid	Sahel Horan	B	I	214
North	Irbid	Tabaqat Fahl	B	0	119
North	Irbid	Wastiyyah	B	0	120
North	Irbid	Al Shoaleh	C	0	120
North	Irbid	Alyarmook Aljadedah	C	0	116
North	Jarash	Jarash Alkubrah	A	I	132
North	Jarash	Alm'arad	B	0	309
North	Jarash	Bab Amman	C	0	169
North	Mafrq	Rhab Aljadedah	B	0	201
North	Mafrq	Sabha & Defianeh	B	I	119
North	Mafrq	Serhan	B	I	141
North	Mafrq	Alza'tary & Almansheah	C	I	139
North	Mafrq	Aum Qutain & Makfieah	C	0	101
North	Mafrq	Dair Alkahf Aljadedah	C	I	110
North	Mafrq	Khaldiyyah	C	I	210
North	Mafrq	Manshiat Bane Hasan	C	0	132
North	Mafrq	Bani Hashem	D	0	88
South	Aqaba	Aqaba	A	0	593
South	Aqaba	Hud Aldisah	C	I	89
South	Aqaba	Wadi Araba	C	I	87
South	Aqaba	Qatar & Rahmah	D	0	55
South	Karak	Abdulah Bin Ruaha	B	0	213
South	Karak	Ayy	B	0	140
South	Karak	Mu'ab Aljadedah	B	0	358
South	Karak	Qatraneh	B	0	117

Region	Governorate	Municipality	Class	Treatment	Sample
South	Karak	Sultani	D	0	118
South	Maan	Shobak Aljadedah	B	0	136
South	Maan	Al Jafer	C	1	145
South	Maan	Alsharah	C	0	120
South	Maan	Iel Jadedah	C	0	195
South	Tafela	Al Tafeilah Alkubrah	A	1	374
South	Tafela	Qadesiah	C	0	228

Sample sizes range widely based on municipality population, ranging from 55 respondents in the Class D municipality of Qatar & Rahmah to 982 respondents in the Class A municipality of Zarqa.

Related to balance tests of the MICA data, the following graphic shows the normalized differences of these items across treatment and comparison municipalities, where standardized differences greater than 0.25 indicate a potential validity threat to the evaluation measures. The following table provides the detail on treatment and comparison measures.

Table 29: Levels and standardized differences, CITIES public perception indicators

Hypothesis	Indicator	Measure	Treatment	Comparison	Difference	Standardized difference
2	Quality services	of Drinking water	2.33	2.36	-0.03	-0.02
2	Quality services	of Irrigation / water maintenance	3.08	2.96	0.12	0.10
2	Quality services	of Sanitation	2.63	2.72	-0.09	-0.08
2	Quality services	of Roads / bridges	2.76	2.46	0.3	0.24
2	Quality services	of Health facilities	2.79	2.87	-0.08	-0.07
2	Quality services	of Schools	3.01	3.03	-0.02	-0.02
2	Quality services	of Electricity	3.45	3.37	0.08	0.09
2	Quality services	of Roads and bridges	2.33	2.51	-0.18	-0.14
2	Quality services	of Public spaces	2.1	2.12	-0.02	-0.02
2	Quality services	of Trash removal	2.74	2.86	-0.12	-0.10
2	Quality services	of Services index	2.74	2.74	0.00	0.00

Hypothesis	Indicator	Measure	Treatment	Comparison	Difference	Standardized difference
2	Quality of services	Services improved over past year	2.81	2.81	0.00	0.00
3	Responsiveness	Local council	2.71	2.58	0.13	0.12
3	Responsiveness	Municipal council	2.61	2.46	0.15	0.13
3	Responsiveness	Governorate council	2.77	2.66	0.11	0.09
4	Effectiveness	Local council	1.94	2.09	-0.15	-0.14
4	Effectiveness	Municipal council	2.08	2.27	-0.19	-0.18
4	Effectiveness	Governorate council	1.88	2.02	-0.14	-0.13
5	Cohesion	Perceived ability to solve common problems	2.59	2.67	-0.08	-0.07
6	Confidence	Local council	2.07	2.16	-0.09	-0.08
6	Confidence	Municipal council	2.15	2.26	-0.11	-0.10
6	Confidence	Governorate council	2.04	2.09	-0.05	-0.04
7	Stability	Direction of municipality	0.39	0.41	-0.02	-0.04
7	Stability	Direction of municipality compared to previous year	2.99	2.99	0.00	0.00
7	Stability	Standardized index of responsiveness, effectiveness, and confidence	-0.09	0.08	0.17	0.18

ANNEX 6: ANALYZING INSTITUTIONAL CAPACITY

One of the key outcome measures for the CITIES IE is municipality MICA score. In developing this evaluation design, the evaluation team met with the CITIES Component One Lead several times to learn about the background and implementation of the MICA tool. It is critical to understand how CITIES is using the MICA tool, what it can and cannot capture, and the potential for bias or omission based on the tool's design.

Originally created as part of a Chemonic's activity in Afghanistan, the MICA tool that CITIES is using has been adjusted for the Jordanian context. The Component One lead described it as a vehicle for, "entering into a first conversation with Municipalities. It gives the program [i.e. CITIES] - despite the admitted methodological flaws of the tool - a picture of where the municipalities are in terms of service delivery."

CITIES is in the process of applying the MICA tool in 33 Municipalities. The tool is applied over the course of two to three days and is used as an interview template with relevant stakeholders. Enumerators do not directly ask about each MICA metric, but rather pose open-ended questions related to each metric and then attempt to substantiate these through follow-up questions and requests to observe documents, certificates, or other relevant materials and fixed assets. After the scoring process in the Municipality is finalized, the MICA enumerators discuss the findings with the Component One lead to collaboratively review and determine a final score for the Municipality.

After a municipality has been scored, CITIES staff develop a service delivery improvement plan. CITIES also identifies which of the four activity tools (grants, technical assistance, training programs, or private public partnerships (PPP)) could be used for implementing the service delivery improvement plan (SDIP). From here, CITIES meets with the mayor, councilors, and community members to prioritize the identified needs. For example, based on the MICA scoring, CITIES may identify solid waste management, urban planning, and fire prevention management as the main areas of need. The municipality stakeholders rank the identified needs that emerge from the MICA based on local priorities. In a final step, the prioritized needs are sent to the Municipal council for approval. In parallel to this, the SDIP may also identify needs through community outreach and to capture priorities that may not be addressed in the MICA.

MICA CHALLENGES

Article 5A of the 2015 Municipality Law states that councils shall assume 29 functions, authorities and powers within the boundaries of the municipality. Only 18 of the 29 functions fall under the full control of the municipalities, while the other eleven occur "in coordination" with other entities, such as the sector ministries, although there is no guidance in the law regarding this coordination.

Of the 23 MICA scoring metrics, six are not under the locus of control for municipalities, as shown in the table below. To

Measuring Beyond Municipality Functions

The analysis of which functions are under the full control of municipalities is based on the land law and the evaluation team's research and interviews. However, in many cases, it is possible that CITIES activities may have an influence on MICA items that are not under the control of municipalities, as detailed in the table below, due to citizen engagement efforts. For example, if local leaders respond to complaints related to areas outside municipal control, they may make an effort to contact the relevant authority higher in the governance structure. The MICA tool may in fact pick up this change should it occur.

complicate matters further, review of the MICA tool, municipal law, and other documentation reveals that not all of the MICA indicators that fall under the direct control of the municipalities are anchored in existing legal, regulatory, and policy frameworks (e.g. citizen service desk is not a legal obligation of the municipalities). There are 14 MICA metrics that both fall under the direct span of control of municipalities and are anchored in the law and regulatory framework. These items are highlighted in blue in Table 30.

Table 30: MICA Items Under Municipal Control

MICA No	Service	Under Control	Municipal
1.1	Citizen services	Yes	
1.2	Manual records & systems for Permit & licensing system	Yes	
1.3	Municipality service orientation toward individual citizens	Yes	
2.1	Technical trainings in service delivery	Yes	
2.2	Municipal Staff capacity for basic administrative and management skills	Yes	
3.1	Urban planning	Yes	
3.2	Capacity to design local development plans and manage design and execution of capital projects	Yes	
3.3	Participation of citizens & business community in identifying citizens' needs	Yes	
3.4	Building licensing and control	Yes	
4.1	Services r.t infrastructure	Yes	
4.2	Maintenance of assets	Yes	
4.3	Solid waste management	Yes	
4.4	Cleanliness of public spaces and infrastructure	Yes	
4.5	Public utilities	No	
4.6	Wastewater management	No	
4.7	Public transport	No	
4.8	Street naming & building numbering	Yes	
5.1	Markets	Yes	
6.1	Tourism, cultural, and leisure time services	No	

MICA No	Service	Under Control	Municipal
6.2	Public health related services	Yes	
7.1	Fire prevention	No	
7.2	Calamity management	No	
8.1	Planning and reporting	Yes	

The MICA tool is likely sufficient for CITIES purposes of “starting a conversation” around municipality service needs. However, to fill the gap left by the current MICA items that are neither under municipal control nor supported through regulatory guidance, the evaluation team will consider employing additional MICA items that capture changes in municipality services, processes, and management. An augmented MICA will allow the evaluation team to better measure outcomes that have the potential to change in both treatment and comparison municipalities and are related to CITIES interventions.

ANNEX 7: LISTING OF SELECTED TREATMENT AND COMPARISON MUNICIPALITIES

Region	Governorate	Municipality	Treatment	Class	Log population	Refugee population	Poverty rate
Central	Amman	Sahab	I	B	11.767	0.195	0.202
Central	Balqa	Dair Alla	I	B	10.841	0.017	0.299
Central	Balqa	M'addi	I	B	10.018	0.049	0.299
Central	Zarqa	El-Hallabat	I	C	10.265	0.602	0.138
North	Irbid	Sahel Horan	I	B	11.184	0.242	0.251
North	Mafraq	Bal'ama Aljadedah	I	B	10.419	0.099	0.144
North	Mafraq	Dair Alkahf Aljadedah	I	C	9.298	0.046	0.428
North	Mafraq	Khaldiyyah	I	C	10.581	0.169	0.195
North	Mafraq	Prince Alhusain Ben Abdollah	I	C	9.750	0.165	0.061
North	Mafraq	Salhiah & Naifeh	I	C	9.939	0.163	0.447
North	Mafraq	Serhan	I	B	10.178	0.239	0.202
North	Mafraq	Um aljmal Aljadedah	I	B	10.289	0.193	0.176
South	Aqaba	Hud Aldisah	I	C	8.719	0.017	0.475
South	Maan	Al Jafer	I	C	8.962	0.004	0.338
South	Maan	Husanieh Aljadedah	I	B	9.814	0.008	0.525
Central	Amman	Jizah	0	B	11.274	0.086	0.222

Region	Governo rate	Municipa lity	Treatme nt	Class	Log populati on	Refugee populati on	Poverty rate
Central	Balqa	Ain Albasha	0	B	11.589	0.075	0.295
North	Ajlun	Alauion	0	B	10.002	0.057	0.339
North	Ajlun	Shafa	0	B	10.128	0.062	0.278
North	Irbid	Al Seru	0	B	10.042	0.119	0.132
North	Irbid	Alyarmoo k Aljadedah	0	C	10.109	0.157	0.132
North	Irbid	Mazar Jadedah	0	B	10.037	0.148	0.170
North	Irbid	Mo'ath bin Jabal	0	B	10.849	0.147	0.360
North	Irbid	Tabaqat Fahl	0	B	10.558	0.180	0.360
North	Jarash	Bab Amman	0	C	9.735	0.026	0.203
North	Mafraq	Husha Aljadedah	0	C	10.131	0.264	0.061
North	Mafraq	Manshiat Bani Hasan	0	C	9.382	0.236	0.150
South	Karak	Talal Aljadedah	0	C	8.970	0.021	0.286
South	Maan	Alsharah	0	C	9.619	0.008	0.505
South	Maan	Iel Jadedah	0	C	9.625	0.018	0.483

ANNEX 8: QUALITATIVE DATA COLLECTION GUIDES

GENERAL INTRODUCTION FOR KIIS/GROUP INTERVIEWS²⁷

Thank you very much for meeting with us today and for being willing to answer our questions. Before we start, let me provide some context for this meeting and explain briefly what we would like to discuss with you and why. As you know, USAID is supporting a wide range of projects and activities carried out by various organizations in Jordan. With the knowledge and consent of the Government, and as part of its ongoing commitment to adapt and learn from its interventions, USAID has asked us, the USAID Jordan Monitoring and Evaluation Support Project, to conduct a study of municipal public services in Jordan. The purpose of this study is to collect information about municipalities in Jordan at different points in time to understand and measure changes in municipal institutional capacity and public service delivery. Part of the study will cover contextual variables affecting local governance.

The findings from this evaluation will help USAID, the Government of Jordan, and the broader donor community better understand the status of municipal services and improve the effectiveness of their respective interventions.

Your answers will be kept confidential; the report that will develop out of this study will not attribute any particular comment to any particular individual, or for that matter to any particular group of respondents. All we are trying to do is merely understand how municipalities are functioning. We will then review everything we have heard and summarize it in a report that will be shared with USAID.

Again, we are very grateful for your willingness to help us as we conduct this study. If you are comfortable with this approach, I am planning to ask you X questions, a few of which entail follow-up questions. You may end the interview at any time. Before we proceed, do you have any questions for us?

KII DISCUSSION GUIDE FOR EXECUTIVE/GOVERNORATE COUNCIL MEMBERS

1. What do you think of the new decentralized system and how has your experience as a new executive/governorate council member been so far?
2. Executive council members only: How has your role changed under the new system?
3. Governorate council members only: What is your role and how are you performing it?
4. How frequent are the meetings of the executive/governorate council?
5. How does the executive council communicate with the governorate council? How many times did the two councils meet before they endorsed this year's budget?
6. Can you please describe the process by which municipal needs are processed at the governorate level? How do the executive/governorate councils prioritize the development needs/projects they receive from the various municipalities? Can you walk me through the sequence of preparation and approval, number and type of participants involved and reference resources used in the process?
7. Are there information bottlenecks that impact this process?
8. How often is the governorate council briefed on budget disbursements for governorates or municipalities?

²⁷ The introduction will be tailored to the various types of respondents targeted for data collection.

9. Do you think this new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
10. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
11. Is there any negative or positive impact for the implementation of the new system?

KII DISCUSSION GUIDE FOR MAYORS

1. What do you think of the new decentralized system and have you seen any changes so far?
2. **For treatment municipalities only:** Can you brief us about CITIES support to your municipality so far? How do you evaluate this support?
3. Have you been approached by your governorate council members yet? If yes, please describe the process and what resulted from it?
4. Do you think the new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
5. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
6. What are the three most important priorities for improving municipal service provision capacities overall?
7. What specific challenges do you encounter while performing your job ((financial management challenges, general management challenges, working environment challenges, logistical support challenges, other)?
8. Can you suggest management practices that are important for the municipality to adopt, to enhance its internal capacity and facilitate its workflow?
9. Can you walk us through the sequence of budget preparation and approval, number and type of participants involved and reference resources used in the process?
10. Will all capital spending decisions be made at the governorate level while operating expenditures are included in your budget approved by MOMA?
11. Do you think the new position of Executive Manager will streamline municipal workflow? How/why not?
12. Has the Executive Manager attended any of the Executive Council meetings so far? How did that go/ Why not?
13. Does tribalism affect your ability to perform your job? If yes, how?

KII DISCUSSION GUIDE FOR MUNICIPAL COUNCIL MEMBERS

1. Can you please describe your role for me?
 - a. What specific functions does the municipal council perform?
 - b. Is your role effective? How about the role of the local council? Why/Why not?
 - c. Do you feel your role overlaps with/complements that of the local council? What about with certain units within the municipality?
 - d. What challenges specific to your role have you encountered since you started?
2. What do you think of the new decentralized system and have you seen any changes yet?
3. How do local councils engage/consult with the community and then communicate public input/feedback to the municipality?

4. How often do municipal councils meet? Do members get compensated for these meetings? Do you generate meeting minutes? Are they publicly available?
5. How can citizens be sure that the municipality is incorporating the input/feedback they provide? Does engagement result in projects the community has collectively agreed on? How does the municipality report back to the community?
6. Do you think this new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
7. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
8. Is there any positive/negative impact for the implementation of the new system?
9. **For treatment municipalities only:** Are you aware of the CITIES project? Can you brief us about CITIES support to your municipality so far? How do you evaluate this support?

KII DISCUSSION GUIDE FOR LOCAL COUNCIL MEMBERS

1. What do you think of the new decentralized system and have you seen any changes yet?
2. Can you please describe your role for me? What specific functions does the local council perform?
3. Do you feel your role overlaps with/complements that of the municipal council? What about with certain units within the municipality?
4. What challenges specific to your role have you encountered since you started?
5. How do local councils engage/consult with the community and then communicate public input/feedback to the municipality?
6. Is your role effective? How about the role of the municipal council? Why/Why not?
7. How often do local councils meet? Do members get compensated for these meetings? Do you generate meeting minutes? Are they publicly available?
8. How frequently are you briefed on the proceedings of municipal council meetings?
9. How can citizens be sure that the municipality is incorporating the input/feedback they provide? Does engagement result in projects the community has collectively agreed on? How does the municipality report back to the community?
10. Do you think this new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
11. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
12. Is there any negative/positive impact for the implementation of the new system?
13. **For treatment municipalities only:** Are you aware of the CITIES project? Can you brief us about CITIES support to your municipality so far? How do you evaluate this support?

KII DISCUSSION GUIDE FOR CIVIL SOCIETY ORGANIZATIONS

1. What do you think of the new decentralized system and have you seen any changes so far?
2. The new municipalities law mandates that the municipality should engage civil society. Is the municipality engaging CSOs? Are local councils reaching out to you? How/why not?
3. In your opinion, how effective are the municipality's community engagement mechanisms? Are they sufficiently inclusive? Does engagement result in projects the community has agreed on?
4. Is the role of the local councils effective? Why/Why not?

5. How do you monitor the performance of your municipality?
6. Do you think the new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
7. Does the municipality engage civil society in the budgetary process?
8. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
9. Is there any negative/positive impact for the implementation of the new decentralized system?
10. **For treatment municipalities only:** Are you aware of the CITIES project? Can you brief us about CITIES support to the municipality so far? How do you evaluate this support?

KII DISCUSSION GUIDE FOR GOVERNMENT OFFICIALS

1. What is your role?
2. There are a number of donor-led projects supporting decentralization and municipal capacity building. Can you speak a little about the coordination effort and structure?
3. Are you aware of the CITIES project?
 - a. Can you brief us about CITIES support to municipalities so far?
 - b. How effective has this support been?
4. What do you think of the new decentralized system?
 - a. Is there any negative/positive impact for the implementation of the new decentralized system?
 - b. What regulations are still needed to streamline the new decentralized process?
 - c. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
5. In your opinion, what sequence of interventions would best support decentralization today given current limitations and political economy realities? What should donors focus on?
6. What interventions would best support public service delivery and municipal capacity building?
7. Because of the current economic situation, fiscal decentralization will be a decentralization of expenditures but not revenue collection. This means municipalities will continue to depend on the central government. Do you agree with this statement? Please elaborate.
8. What is the formula for deciding governorate budgetary allocations?
9. I am aware that there are different interventions that aim to build the capacity of the various councils through training. How sustainable will this learning be if/when councilors change in the next election? Do you think it is better to focus capacity building efforts on municipal staff? Please elaborate.
10. Will the new Local Governance Institute be training council members and municipal staff? What skills will you be focusing on? If training is meant to cover the roles of the various councils, will it also tackle technical competencies to perform these roles?

KII DISCUSSION GUIDE FOR DONORS/USAID IPS

1. Can you tell me about your project and how it relates to municipal capacity building, public service delivery, and/or decentralization?
2. Has the CITIES program coordinated or collaborated with you in anyway? Please explain.
3. What do you think of the new decentralized system and have you seen any changes so far?

- a. Do you think the new decentralized system will improve the quality of public services delivered at the municipal level? Why/Why not?
 - b. Is there any negative/positive impact for the implementation of the new decentralized system?
4. What challenges/opportunities do you foresee in the implementation of the new decentralized system?
 - a. What regulations are still needed to streamline the new decentralized process?
5. In your opinion, who will be the emerging power holders in the new system, and what is the nature of their respective power?
6. In your opinion, what sequence of interventions would best support decentralization today given current limitations and political economy realities?
7. In your opinion, what sequence of interventions would best support municipal service delivery and capacity building today, given current limitations and political economy realities?

KII DISCUSSION GUIDE FOR CITIES

1. Can you please update us about any progress in your component/generally?
2. Have there been any changes on the ground affecting the implementation of the project?
3. Is the planned sequence of interventions proving to be sufficiently suited to the context? Please elaborate.
4. What challenges/opportunities, if any, have your component/project encountered?
5. Are there any outliers in the treatment municipalities? If yes, what factors contribute to this designation?
6. How do you evaluate the implementation of the new decentralized system? Is there any negative/positive impact so far?
7. What regulations are still needed to streamline the new decentralized process and support the work you are doing?
8. Are the municipalities and/or the GoJ sufficiently cooperative?
9. What is your current mechanism for coordination with other donors and/or the GoJ?
 - a. With whom do you coordinate within these organizations/institutions?
10. Do you think planned capacity building support to various council members and municipal staff will be able to bridge current technical gaps? Please elaborate.
11. Do you think the decision to house the local governance institute at the Amman municipality will contribute to the sustainability of the project's capacity building efforts? If yes, how? If no, why not?

KII DISCUSSION GUIDE FOR USAID

1. CITIES has been rolling out various activities at the municipal level. What do you think of the progress the project is achieving so far?
 - a. Which interventions do you view are more effective? Less effective? Why?
 - b. Are there already any outliers in the treatment municipalities? If yes, what factors contribute to this designation?
 - c. Is the planned sequence of interventions proving to be sufficiently suited to the context? Please elaborate.
2. What challenges/opportunities if any has the project encountered so far?

- a. Have there been any changes on the ground affecting the implementation of CITIES?
 - b. Have any lessons learned emerged thus far?
- 3. What types of collaboration or coordination are you aware of between CITIES and other USAID IPs or other donors?
- 4. Are the municipalities and/or the GoJ sufficiently cooperative?
- 5. Do you think planned capacity building support to various council members and municipal staff will be able to bridge current technical gaps? Please elaborate.
- 6. Do you think the decision to house the local governance institute at the Amman municipality will contribute to the sustainability of the project's capacity building efforts? If yes, how? If no, why not?
- 7. How do you evaluate the implementation of the new decentralized system? Is there any negative/positive impact so far?
 - a. In your opinion, what regulations are still needed to streamline the new decentralized process and support the work CITIES is doing?

ANNEX 9: RELEVANT LEARNING AGENDA FINDINGS

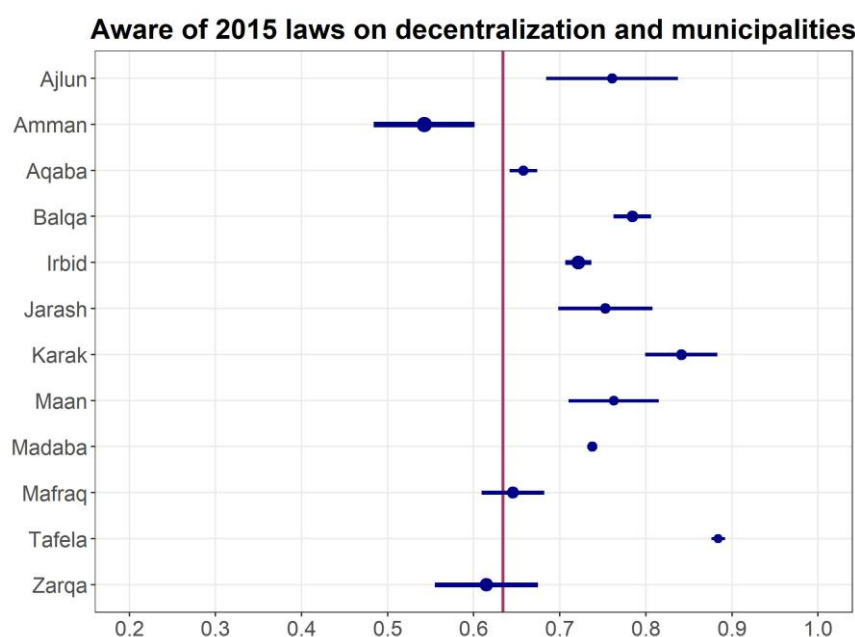
Two of the learning agenda themes are directly relevant to the CITIES IE:

- Democracy, human, rights, governance
 - This theme addresses questions of citizen awareness, decentralization, and engagement.
- Fragility/resilience
 - This theme incorporates findings around refugees and migrants.

DEMOCRACY, HUMAN RIGHTS, AND GOVERNANCE

What is the level of citizen awareness and knowledge of Jordan's decentralization agenda? At what level of governance does subsidiarity [decentralization] most effectively reside—local councils, municipal councils, or governorate councils? What is the locus of control for all three levels of government?

The level of citizen awareness and knowledge of Jordan's decentralization agenda is somewhat high among certain groups, but confusion around subsidiarity is also high. Around 64 percent of Jordanians are aware of the decentralization and municipalities laws. The distribution of this knowledge is not equal across the country, with only 55 percent of people in Amman expressing awareness of decentralization (the lowest rate), while citizen awareness in Tafileh governorate is around 88 percent (the highest rate).

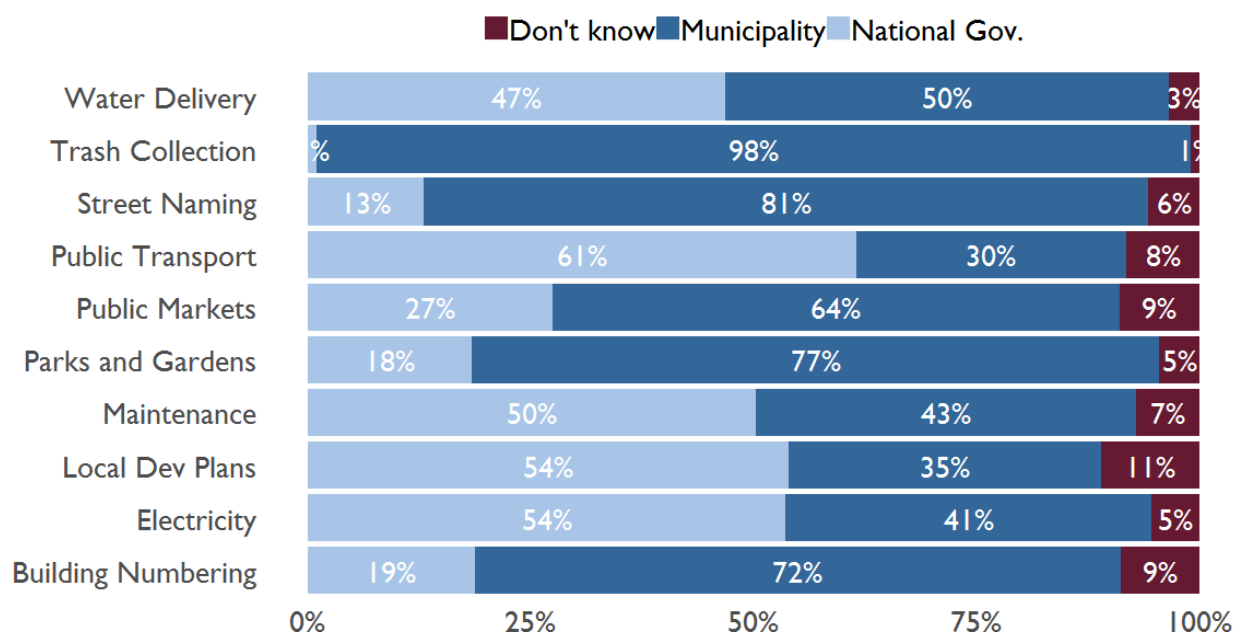


People with more education reported a higher rate of decentralization awareness, with 83 percent of master's degree holders reporting awareness of the relevant laws, while 46 percent of respondents with an elementary school education were aware. Respondents who voted in the 2017 elections reported a higher awareness of decentralization: 77 percent of those who voted were aware of the two decentralization-related laws compared to 58 percent of those who did not vote. Across all education levels, awareness of decentralization was higher among voters than non-voters; for example, 95 percent

of voters with master's degrees were aware of Jordan's decentralization agenda, while 74 percent of non-voters with the same level of education were aware of the agenda.

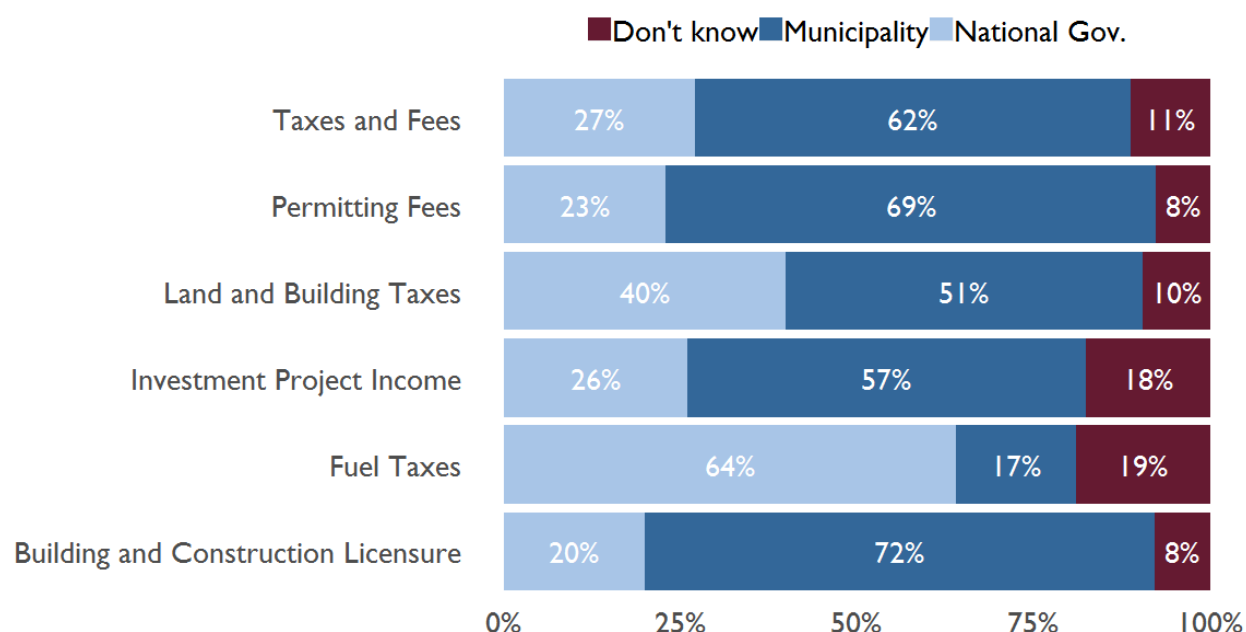
There are two key areas for determining where decentralization is perceived to reside: government functions and government revenue. In four areas—public transportation, maintenance of local assets, local development planning, and electricity—a majority of people thought the locus of control is with the national government. In six out of ten domains, Jordanians believe their local municipality is primarily responsible for services. The case of electricity is notable in that 46 percent of respondents think their local municipality is responsible or do not know who is responsible, despite this falling under the purview of national service provision. The figure below highlights areas where additional outreach may be needed to inform citizen about government roles and responsibilities.

FIGURE 42: PERCEPTIONS OF LOCUS OF CONTROL FOR SERVICE PROVISION RESPONSIBILITY BY DOMAIN

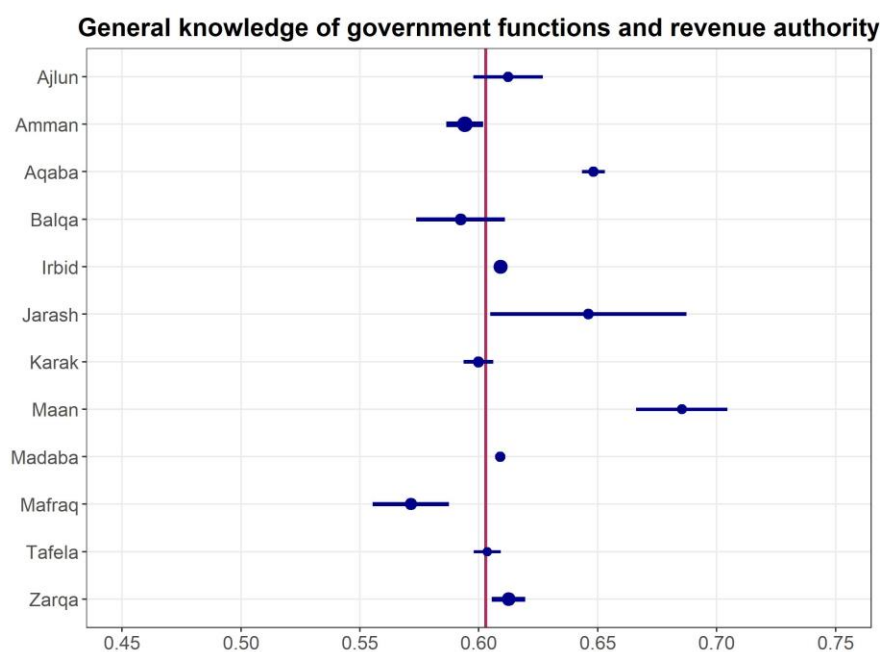


There is more uncertainty around the locus of control for revenue collection, with about one in five respondents unsure about who is responsible for collecting fuel taxes. As shown in the figure below, at least a fifth of respondents across six revenue sources reported that the national government is legally responsible for collecting funds.

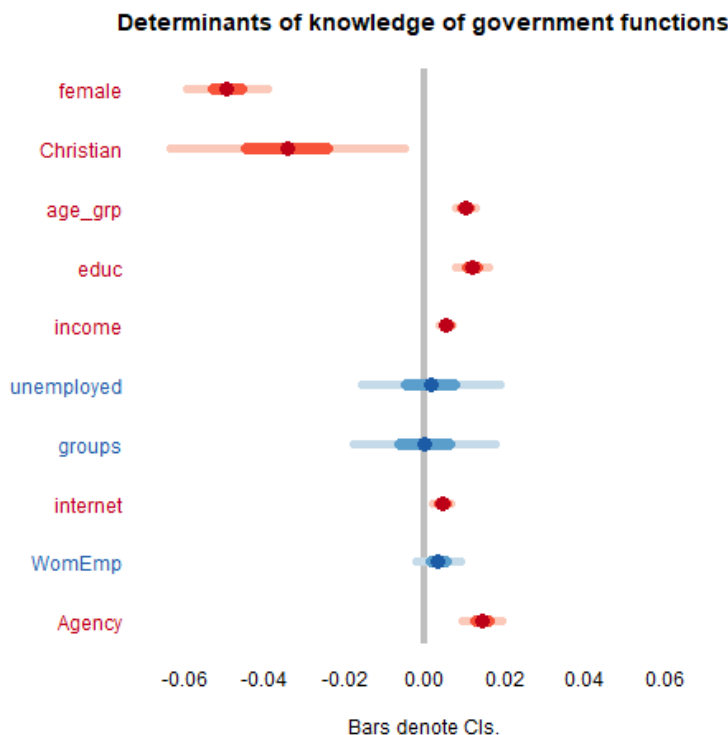
FIGURE 43: PERCEPTIONS OF RESPONSIBILITY FOR COLLECTING REVENUE



When the above analysis is categorized by correct or incorrect knowledge about local government functions and revenue authorities based on relevant law, the 16 survey items can be aggregated into an overall knowledge score. On average, respondents correctly answered 9.7 of 16 knowledge items, or a mean of 60 percent. This score is slightly less than overall awareness of the decentralization and local governance laws (64 percent), and in both cases Amman scores lower than other governorates.



Among predictors of knowledge of government functions, respondents who were female or Christian typically scored less on the knowledge measures. Knowledge was increasing in age, education, income, internet use, and feelings of personal agency and empowerment.

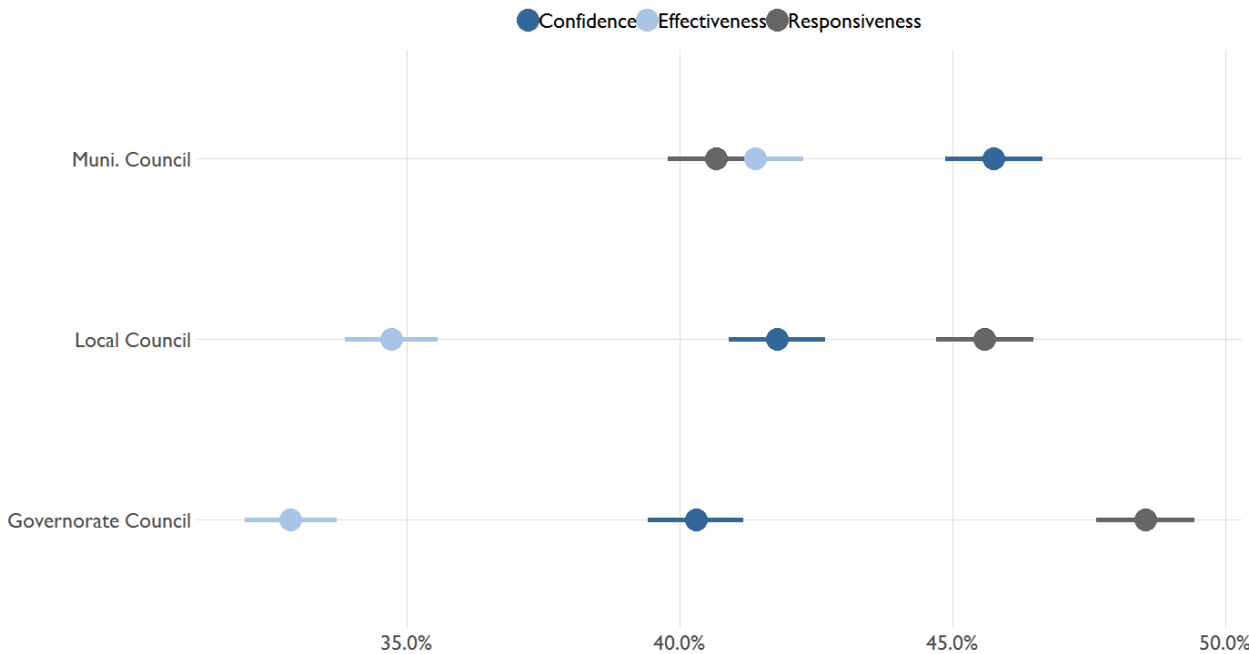


Knowledge of local government functions is not a strong predictor of governance indicators responsiveness, effectiveness, and confidence.

What is the relationship between citizen engagement and citizen participation? Does stronger engagement and participation lead to improved government effectiveness and legitimacy?

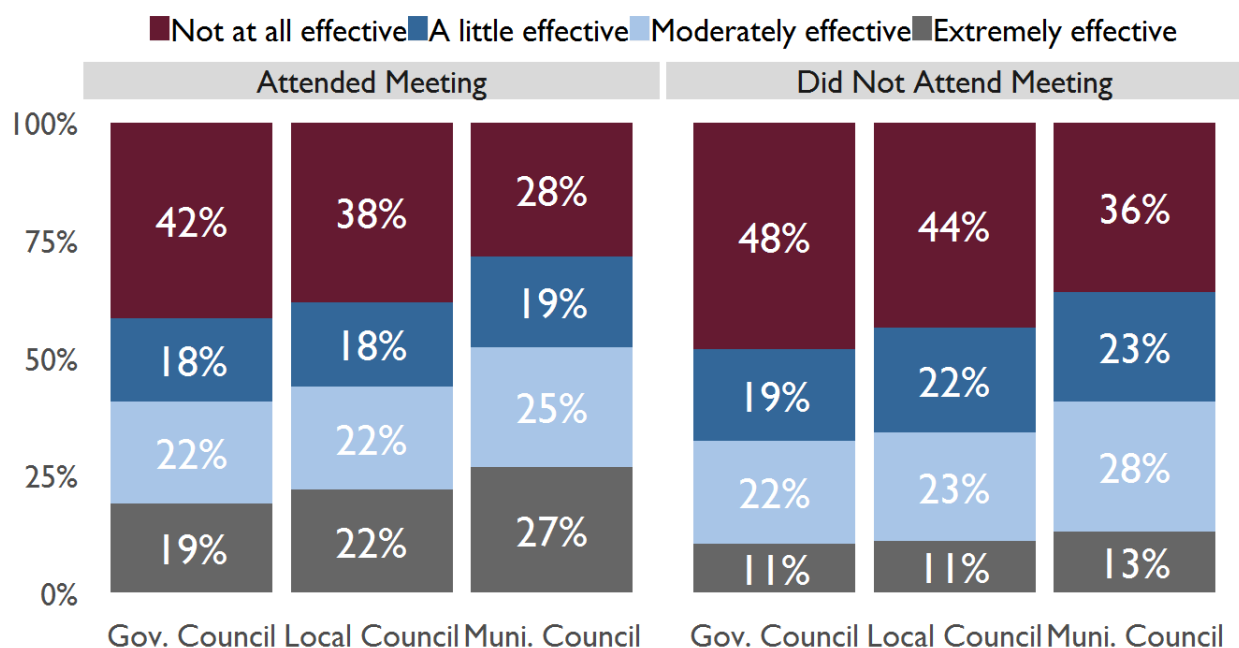
Citizen engagement and participation is generally low, but does appear to affect perceptions of effectiveness, with respondents who have made visits to their local officials perceiving local government as less effective. The relationship between perceptions of effectiveness and participation is complex but does not vary across levels of government. The data suggest that levels of citizen perception vary on a spectrum from responsiveness of local institutions to confidence in these same institutions.

FIGURE 44: PERCENTAGE OF PEOPLE WHO FEEL THEIR LOCAL GOVERNANCE INSTITUTIONS ARE RESPONSIVE, EFFECTIVE, AND HAVE CONFIDENT IN THE INSTITUTION



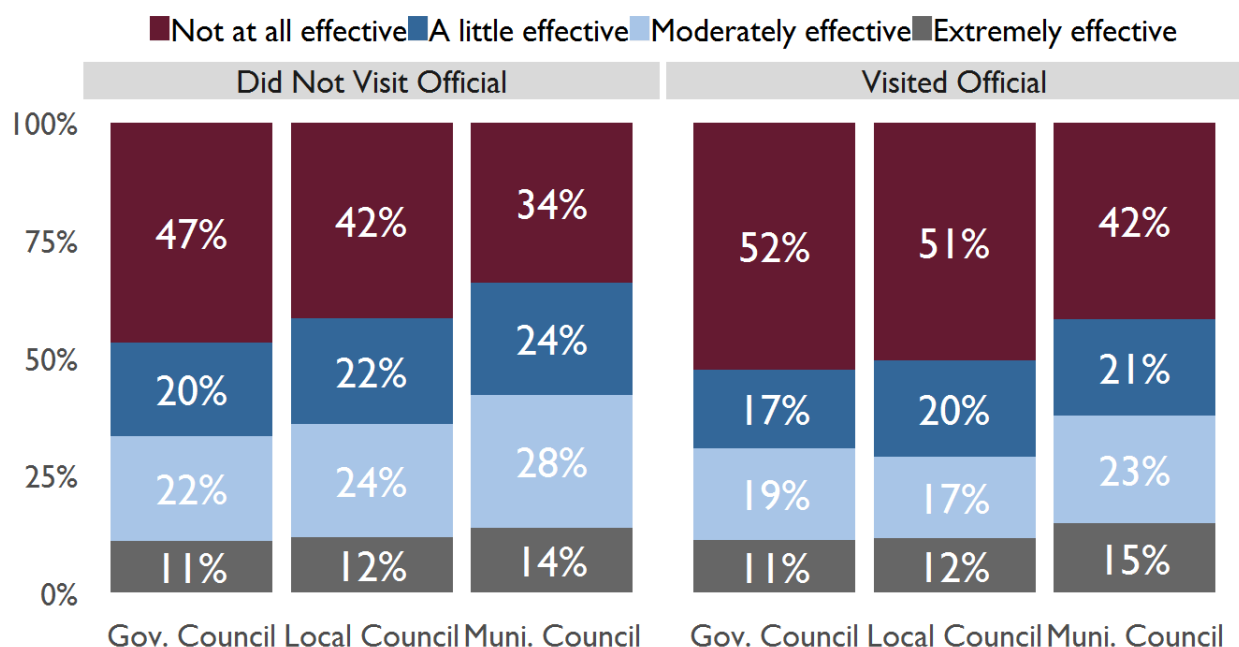
Perceptions differ based on the level of respondent engagement. As shown in the figure above, descriptively, meeting attendance appears to increase the rate at which people view their councils as “extremely effective.” This effect is most obvious at the municipality level, where 27 percent of respondents who attended a meeting in the past year said the municipality council was extremely effective compared to 13 percent of those who did not attend a meeting.

FIGURE 45: PERCEPTIONS OF EFFECTIVENESS BY MEETING ATTENDANCE



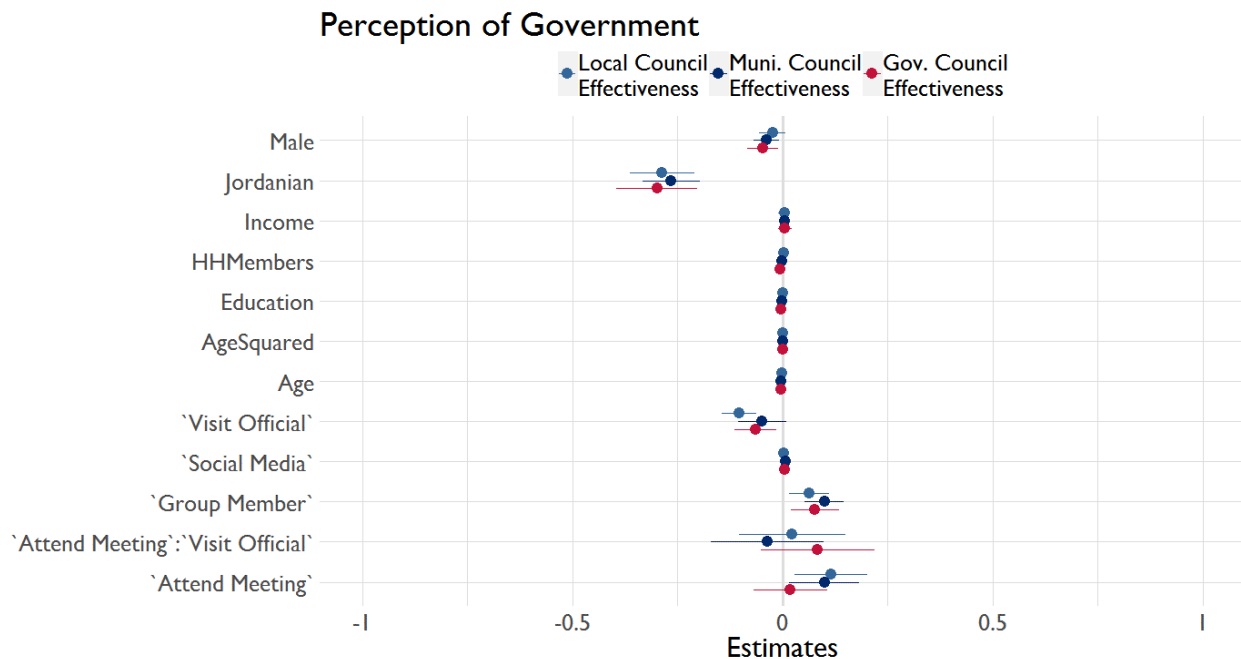
To better understand the relationship between participation and perceptions of effectiveness, we use a model similar to that in the employment section, controlling for key demographic and geographic features. The effect of meeting attendance on perceptions of local government does appear to exist in the descriptive data, as shown above. To understand the strength of support, we augment the original model with a “dosage” effect for visiting a local official. The interaction of meeting attendance and visiting an official provides a measure of the severity of local engagement. This is based on the assumption that more engaged citizens visit their local government officials. Respondents who reported visiting a local official in the past 12 months reported that the government at all levels—local, municipality, and governorate—are not at all effective more frequently. For example, more respondents who visited an official said their governorate council was not at all effective than any of the other response options combined. It may be the case that visiting local officials actually has a negative impact on perceptions of government. Indeed, across all levels of government there is a small and negative correlation between perceptions of effectiveness and visitation with officials.

FIGURE 46: PERCEPTIONS OF EFFECTIVENESS BY WHETHER RESPONDENTS VISITED A GOVERNMENT OFFICIAL IN THE PAST 12 MONTHS



For both visiting officials and those attending meetings, more than a third and almost half of respondents tend to view their local government representatives as ineffective. To understand perceptions of effectiveness, a model was created to see what variables affect perceptions of local effectiveness at multiple levels of governance. People who attended public meetings perceive government at the local and municipality levels as more effective compared to those who did not attend a meeting in the previous 12 months controlling for various demographic features and other metrics, such as social media use. This model is consistent with disaggregated measures of effectiveness for each level, as shown in the figure below.

FIGURE 47: MODEL RESULTS SHOWING THE EFFECT OF ENGAGEMENT AND PERCEPTIONS OF EFFECTIVENESS



Meeting attendance did not have a statistically significant relationship with perceptions of governorate councils, although there was still a positive relationship at this level of governance. The relationship between visiting an official and having a perception of effectiveness was consistently negative. Respondents who visited an official were less likely than those who had not done so to perceive their local council as effective. Put simply, there appears to be a strong relationship between making visits to local officials and perceptions of effectiveness, even after controlling for other factors.

The combination of engagement and participation, i.e. visiting an official and attending a meeting, does not appear to have any statistically significant effect on perceptions of effectiveness. It may be that other measures of engagement frequency or severity are more relevant for determining whether more engagement affects perceptions. Alternative model specifications that combined meeting attendance with volunteering, social media use, and group membership, respectively, yielded similarly insignificant results.

A key takeaway from this analysis is that meeting attendance may help improve perceptions of local and municipality councils more so than visits with local officials, social media engagement, or other forms of outreach. The combination of engagement mechanism to strengthen this relationship also does not appear to actually improve perceptions. A larger percentage of respondents who attended meetings viewed councils at all local levels as extremely effective and, controlling for key metrics as well as other forms of engagement, suggests that meeting attendance almost doubles the odds of viewing the government as effective.

FRAGILITY / RESILIENCE

Are migrant population flows disrupting local governance institutions and economic livelihoods?

View of migrants appear to have negligible effect on perceptions of local governance institutions and economic livelihoods, although people seem to have slightly less confidence in their local councils if they believe migrants are having a negative effect on their community. Respondents in the General Population survey do not have a positive view of refugees. Around 64 percent of the population on average views the effect of refugees in their municipality to be, “very negative,” with only about 3 percent of the population in Jerash, Madaba, Tafileh, and Zarqa reporting that the arrival of Syrians or other migrants and refugees has been somewhat or very positive. Of those who say the influx of refugees has been negative (very or somewhat), 90 percent on average reported that the effect on their municipality over the past 12 months is getting worse.

Respondents’ opinions of government services improving, worsening, or not changing over the past year do not appear to have much of a relationship with their views of migrants and refugees. As shown in the figure below, regardless of perceptions of local service provision, more than 60 percent of respondents felt the influx of migrants was, “very negative.” Those who felt local services had worsened a lot in the past year reported a lower rate of “somewhat negative” effects from recent migrants compared to those who reported local services improving a lot, 7 percent compared to 11 percent. Other measures of service provision reflect this trend: respondents’ antipathy toward refugees and migrants transcends perceptions of local governance institutions. For example, 82 percent of respondents who reported no confidence in their municipality council said the effect of migrants and refugees in their municipality was negative, while 72 percent of respondents with a great deal of confidence felt similarly toward migrants and refugees.

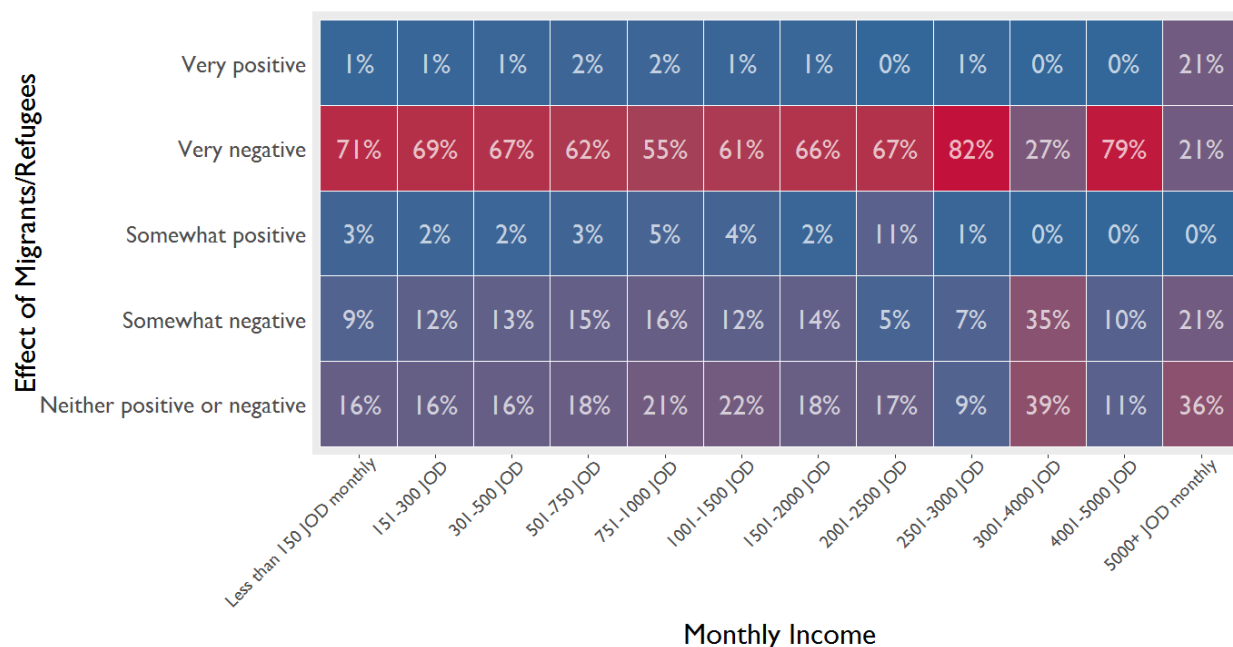
FIGURE 48: PERCEPTION OF GOVERNMENT SERVICES AND PERCEPTIONS OF THE EFFECT OF RECENT MIGRANTS



The perceptions of refugees appear to hold steady regardless of age, employment status, or education level, except for people with master’s and bachelor’s degrees. Around 45 percent of master’s degree holders who feel services have worsened a lot perceive the effect of migrants and refugees as very negative, while 55 percent of bachelor’s degree holders who feel similarly about local services perceive migrants and refugees very negatively.

Respondents with higher monthly incomes reported more positive perceptions of migrants and refugees. Among Jordanians at the highest end of the economic spectrum, earning more than 5000 JOD a month, 21 percent viewed the effect of migrants and refugees as very positive. In the figure below, the darker red suggests a higher density of respondents, while blue suggests a lower density of respondents. A majority of respondents across all income groups, except the very highest earners and those earning between 3000 and 4000 JOD a month, view the effect of migrants very negatively. Notably, these two income groups—between 3000 and 4000 and more than 5000 JOD a month—have generally moderate views of migrants, with a plurality seeing them as neither positively nor negatively.

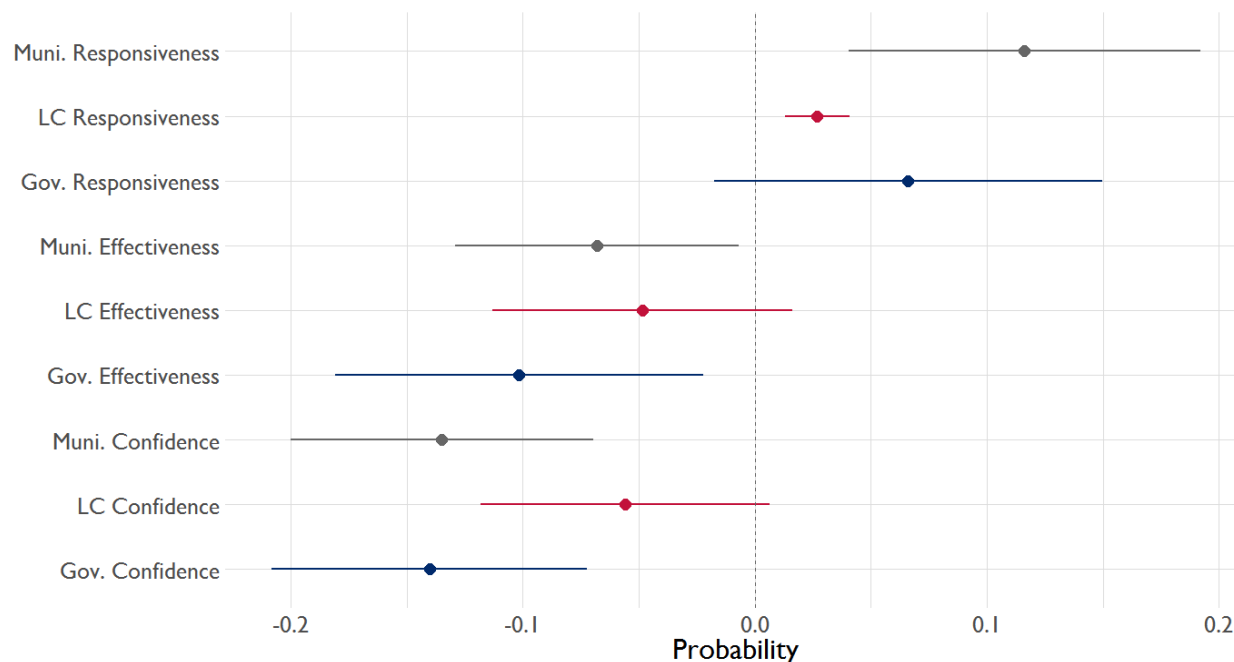
FIGURE 49: PERCEPTIONS OF MIGRANTS BY INCOME GROUPS



Around 74 percent of business owners view the effects of migrants and refugees negatively, while 77 percent of non-entrepreneurs hold the same views. Perceptions of economic opportunity also appear to hold steady regardless of perceptions of migrants. While the perceived constraints to entrepreneurship vary, there does not appear to be any relationship between reported feasibility to start a business and perceptions of migrants and refugees.

Inferential analysis also shows a dynamic relationship between perceptions of local governance institutions. At the local council, municipality, and governorate level, there appears to be a positive relationship between perceptions of refugees and migrants and views of institutional responsiveness, controlling for other metrics such as age, education, and location. The reported effects of migrants and refugees has a negative relationship with how respondents view the effectiveness of local institutions and an even stronger negative relationship with their confidence in these institutions, particularly at the governorate-level. As shown in the figure below,

FIGURE 50: RELATIONSHIP BETWEEN VIEWS OF MIGRANTS AND PERCEPTIONS OF GOVERNANCE INSTITUTIONS; VALUE ABOVE 0 SUGGEST A POSITIVE RELATIONSHIP

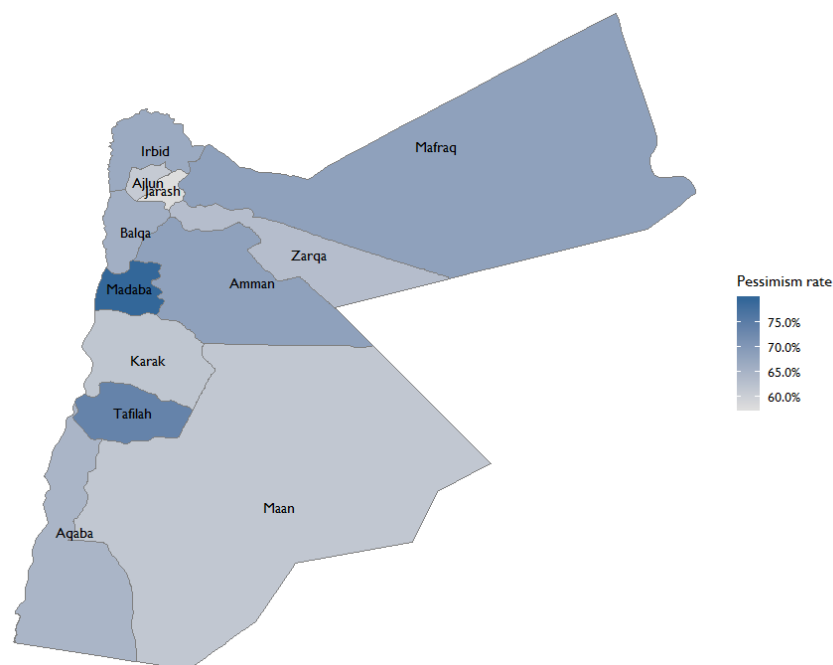


How are local institutions, such as mosques, local meeting groups, CSOs, adapting and responding to the pressures of migrant population flows?

Respondents involved in local institutions appear to be pessimistic about the on-going effect of migrant population flows. The general population survey asked respondents to reflect on how they feel things are changing in their community as a result of the arrival of Syrians and other migrants. Although the survey did not specifically seek out the leaders of local institutions and a single round cannot capture changes in perceptions of these institutions, we can see how participants and non-participants across local groups perceive the effect of migrants on their community.

It could be the case that people are generally pessimistic, so views of migrants and refugees are simply a function of this generally feeling. However, when asked about how things have generally changed in the past year, around 26 percent of respondents said they were getting worse. When asked about the effect of refugees and migrants specifically over the past year, 66 percent of respondents said their effect on the community is getting worse. The distribution of this pessimism about the effect of migrants and refugees is not consistent across Jordan. Respondents closer to Iraq and Syria report that things are getting worse at a lower rate than those in governorates that do not border these countries. Of respondents who were not part of a local community group, 68 percent said the effect of refugees and migrants is getting worse, while 61 percent of those who are part of a local group said the same.

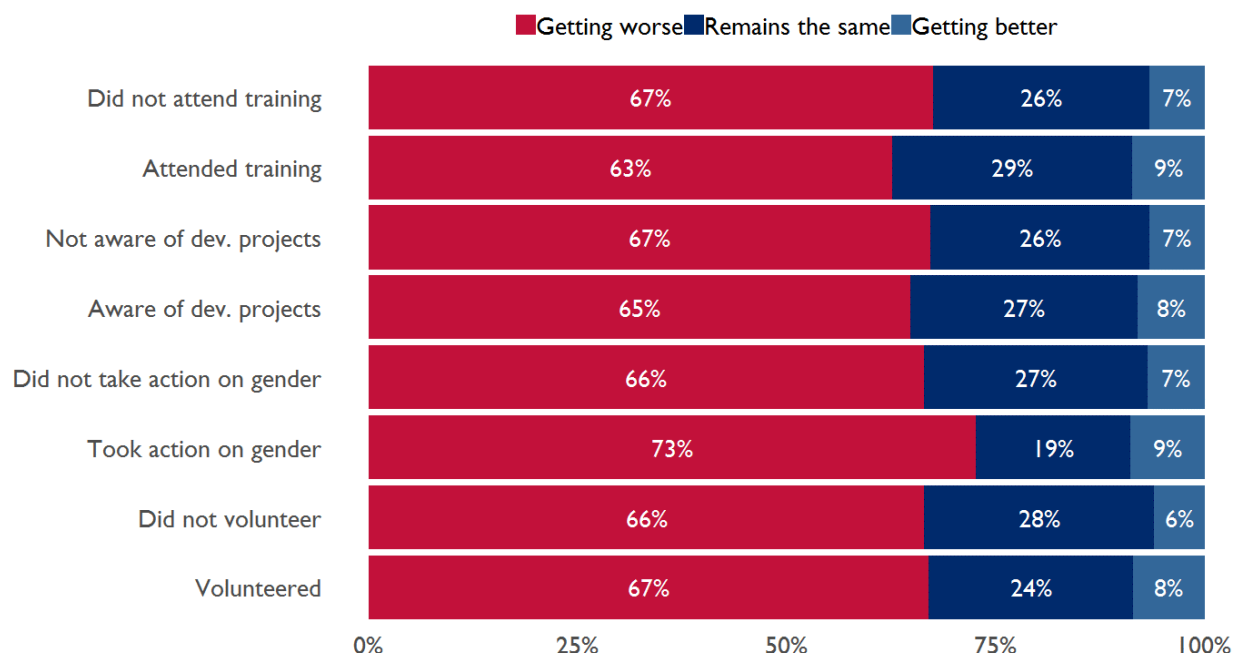
FIGURE 51: OVERALL PERCEPTION THAT THE EFFECT OF MIGRANTS IS GETTING WORSE BY GOVERNORATE



There is some geographic variation within these perceptions, particularly among those involved in local institutions. For example, 72 percent of respondents in Ajloun, who are members of a local group, said the effect of migrants is getting worse, compared to 60 percent of people who are not members of a local group. In contrast, 55 percent of the respondents in Zarqa, who are group members, said the effect of migrants is getting worse compared to 64 percent of non-group members.

Similar to the previous section, respondents generally reported negative feelings about the situation in their municipality compared to the previous year as a result of the arrival of Syrians or other migrants and refugees. Respondents who took action on gender (i.e. through the creation of a gender-related CSO or community group) expressed the highest level of pessimism about the effect of migrants and refugees in their municipalities, with 73 percent saying the effect of migrants is getting worse compared to a year ago. Aside from involvement in local gender groups, descriptive statistics do not reveal a large difference in perceptions among local institution participants and non-participants.

FIGURE 52: PERCEPTIONS OF CHANGE DUE TO MIGRANT INFLOWS AND INVOLVEMENT IN LOCAL INSTITUTIONS



The relationship between survey measures of local involvement and perceptions of how migrant inflows are affecting respondents' local communities appears weak for most local involvement, with the exception of joining a gender-related group. Controlling for education, municipality, age, perceptions of changes over the past year generally, and other contextual metrics, respondents who feel things are getting worse in their municipality as a result of migrants and refugees are about 57 percent on average more likely to join other community members to take action on a gender-related issue (CI 1.13±2.19). This may suggest that in some cases a response among those who feel that the effect of migrants and refugees is getting worse is to form or join gender-related groups. Although this relationship holds over multiple specifications, it is important to note that this is an observed correlation and not a causal association. Indeed, there appears to be no relationship between involvement in other local institutions, such as trainings or volunteering, and a feeling that the effect of migrants and refugees is getting worse, controlling for other contextual factors. For example, local group membership (non-gender related) does not appear to have any meaningful relationship with perceptions of the on-going effect of migrants and refugees. There may be additional analyses that can further investigate the correlation between gender-related groups and perceptions of the effect of refugees and migrants over the past year.

**U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT
1300 PENNSYLVANIA AVENUE, NW
WASHINGTON, DC 20523**