Capacity Maturity and Recommendation

International Civic and Citizenship Education Study (ICPSR 37147)

Simon Chen, Wenjie Tang, Yunwei Zhang, Zecheng Chang

Teachers College, Columbia University

HUDK 4050: Managing Education Data

Professor Amy Nurnberger

May 9, 2022

Abstract

In this article, our group assessed a study done by the International Civic and Citizenship Education Study to examine students' citizenship preparedness with the CMM model (Qin, 2014). We looked over the study's goal and assumptions, data produced, potential challenges, and suggestions, and evaluated the study by each CMM model. Overall, the study records are rather well organized and reusable. The background of the study and the purpose of procedures and variables are well explained. Along with well-recorded questionnaires for all data, the ICCE study got an overall high score between 4 to 5 with our CMM assessment.

Introduction

The International Civic and Citizenship Education Study(2016) is an international assessment aimed to measure to what extent young people from all across the world are prepared to accept their roles and responsibilities when they step into society. Not only did it investigate the study of the students, but it researched and collected the understanding of being a citizen inside a society and students' perspectives and attitudes towards citizenship. For this project, the ICCE study took a look at students from 24 countries on their preparedness for their citizenship. Using civic tests and questionnaires, the assessment is very neatly developed and will be supplemented with lots of information with regards to national and regional information. The researchers looked at data from both the students and teachers to get round perspectives of students' citizenship behaviors in each region. Our team would analyze the project with the CMM model (Qin, 2014) that comes with the instruction and evaluate how much the project fulfills the expectations.

To analyze this study, we would consider our previous reading regarding the Information Security Charter (Information Security Charter, 2014b) and Data Classification (Data Classification, 2014a) when evaluating the study. We would pay attention to the International Association for the Evaluation of Educational Achievement's treatment of sensitive data and confidential data, and how they embedded data security into policies. We would also pay attention to how the organization treats internal and public data, enforces user privacy and data security, and how they corporate with participants to work with data ownership, system ownership, and technical responsibility.

Goal & Assumptions

The study purposes of this project is to address the 4 main questions:

1. International Civic and Citizenship Education Study (ICCS) researched students' knowledge and understanding of civics and citizenship and the factors associated with variations in this civic knowledge. The data was collected by questionnaires and tests to give the feedback of a country's overall level.
2. It gives students' current and expected future involvements in civic-related activities. From these tests, students' perceptions of their capacity to engage in civic activities and their perceptions of the values of attending civic engagement can be measured.
3. Researchers want to find out the students' beliefs about civil and civic issues in society. Also it will give some feedback of concerns with civic organizations, rules, and basic social principles, like democracy and diversity. The data can also tell the students' perceptions of their local communities and threats to the future.
4. It gathers information about the ways of countries organizing civic and citizenship education with focus on the general approaches, the curriculum and its delivery. The processes will benefit the future citizens' civic engagement and interactions in the future.

Based on the four main research questions and goals, the team assembled a document that will analyze the civic projects according to the capability maturity model, and then assess the project with recommendations for potential failures during the process. This project is an international large-scaled assessment, so the process of data collection and processing costs tremendous efforts. We need to make some assumptions that the data collected is true and can

represent the testers' original wills. Also the distribution of the education system is balanced so that the sampling strategies could be effective as expected.

Data Produced

The data was collected with Probability Proportional to Size (PPS) sampling strategies in 2015 - 2016 under the auspices of the International Association for the Evaluation of Educational Achievement (IEA). ICCS collected data from more than 94,000 eighth grade (or equivalent) students in about 3,800 schools from 24 countries. These student data were augmented with data from more than 37,000 teachers in those schools and further contextual data collected from school principals and national research centers.

There are totally 139 datasets with different topics. In the questionnaires, they collected useful survey data:

Datasets 1-24 correspond with the school questionnaire.

Datasets 25-48 correspond with the International Student Achievement Questionnaire.

Datasets 49-63 correspond with the European Module Student Questionnaire.

Datasets 64-87 correspond with the International Student Questionnaire.

Datasets 88-92 correspond with the Latin American Module Student Questionnaire.

Datasets 93-116 correspond with the International Student Reliability Questionnaire.

Datasets 117-138 correspond with the Teacher Questionnaire.

Dataset 139 corresponds with the National Context Questionnaire. (ICPSR37147, 2018)

The public-use data files in this collection are available for access by the general public. Access does not require affiliation with an ICPSR member institution. ICPSR currently

maintains five copies of its data (and requires that any off-site backup be encrypted): One copy held onsite on network-attached storage (NAS) One copy put on tape once a month onsite, with each copy held for 12 months.

Potential Challenges & Recommendations

The biggest challenge is during the survey designing procedure. Researchers need to describe the relevance and procedure to develop a theoretical framework to undergird such studies. In this case, we recommend that the research team could depict potential major operational challenges in advances, including procedures to integrate state-of-the-art theory, relevance for policies of participating countries, cultural relevance of measures, data comparability and the current expansion of these studies to low- and middle-income countries. (Van de Vijver et al., 2019)

The challenges also exist due to working in multidisciplinary scientific teams and working with stakeholders that have a policy perspective on ILSAs. We recommended that the cooperation protocols and standards need to be set up specifically to clarify the procedures and outcomes.

Legal and Ethical Concerns

The project collects data from eighth-grade students and teachers from 24 countries. Student data, especially for children under certified adult age is considered very sensitive and is expected to be treated with great caution. The legal age for each country, like Korea compared to the rest of the world, could be different. And the legal restrictions on student information access could also vary. Though some of the differences are mentioned like the age difference among

countries, the researchers did not particularly discuss their legal responsibilities when coming up with measures, instead focused more on data accuracy and dropped all the details for collected data and measures.

The lack of explanation could come from the fact that those researchers have had prior working experience with most of the schools during 2009. But we would recommend the researchers share more insights in every project regarding legal and ethical considerations when they come up with measures and variables.

Sampling & Data Collection

As collecting data from 24 countries, sampling and the data collecting process could be a challenge. Different countries, regions, and school districts would be varied in resources and access. Translating questions and responses with accurate meaning in different languages and communicating with the internal group to record the data could be a challenge. The available file types for data collection could also vary in different regions. And being able to randomly select schools and students that best represent each country and culture is nonetheless one of the hardest challenges.

The ICCE study chose to continue working with the schools that they had established prior connections with. Though those schools' selections may not be random, the overall process for data collection may be more cost-efficient and ensure better data quality. The data file used for questionnaires is used as SPSS and SAS. And all the variables, data records sources, questionnaires, and scores evaluations are explained with details on usage and purpose.

Overall, we believe the researchers did a rather good job documenting all the measurements and data records. But we would recommend that they address more details

regarding the data collection process, how raw data was stored and translated into organized data, and how data are passed on from different regions and organized in internal communication. We would like to learn more about the privacy considerations practiced during the data collection process. And whether the community faced any challenges in protecting students' and teachers' data.

Regarding data quality, we believe that the potential for making more diverse and grounded questionnaires to define civic manners based on culture is still apparent. The definition of civic could be varied in culture and in different situations. Using the same set of questionnaires to evaluate both USA and Japanese students' civic behavior could be biased. We are aware that requesting researchers to come out with appropriate questionnaires for each culture could be too ambitious and cost effort. But we still suggest that researchers could keep diversity in mind when selecting questionnaires and translate questions considering their ground environment.

Storage and Privacy Concerns

As we mentioned earlier, we encourage the researcher to be more transparent with internal data sharing and communication. The researchers did a very good job recording all the organized data files and keeping readable and reusable data records on ICPSR. The ICPSR download tracking allows the researchers to keep track of the use of their data and provide tracking if the data became misused. But outside of ICPSR, we don't know how the researchers keep the raw data and how all the data are maintained. We would recommend the researchers explain more details about how the data was maintained outside of the ICPSR, how privacy was

considered during internal communication and external data sharing, and saving data backups for possible data loss.

Data Sharing and reusability

The researchers include preferred citations with the record to guide legal citations. Saving the record on ICPSR allows the organization to keep track of download and usage reports for the potential use of the study to stay alert of potentially harmful use of data. Regarding the overall data records, the purpose of the studies, the records of data sources and measures, the purpose of variables and mechanics, and all the questionnaire files are all clearly defined and saved. The overall data records are very well organized making the study very much reusable and sharable. Our only suggestion would be to encourage the research team to communicate with each school community prior to and during the data collection process. Communicating with the data wonder allows the researchers to make sure their plan is practical under local restrictions, and discuss how to make the data sharable and beneficial to the local community.

Project Analysis per Capability Maturity Model

The capability Maturity Model (CMM) is a guideline or framework that can help us to evaluate and assess the data management capability of a project. Since artificial intelligence starts playing a very important role in almost every industry, and along with the tremendous increase in data processing speed, more and more companies and organizations start paying more attention to data management plans, because the data that a company or organization creates is a very valuable asset and resource to them. Just having a Data Management Plan is not enough, we still need to evaluate it in case the data management plan is mature enough to support the project,

and here comes the Capability Maturity Model. There are 5 rubrics within the model and 6 levels for each of the rubrics scored from 0 to 5. Overall, this project is doing very well with its data management plan. They have a technical report that includes a completed data management plan, we are going to evaluate this data management using this 5 rubrics one by one.

1. Data Management in General

For the first rubric, the overall goal is to make sure that the data management plan was designed aiming to collect and maintain high-quality data. There are 15 sub-rubrics we evaluated and this project is doing well in this category. However, this project didn't talk about how they developed and implemented a budget while doing this project. We believe that they must have this information and they just didn't publicize it. So, we mark this sub-rubric as a 0 and others are all 5, and the average ends up with a 4.6 out of 5, which makes this project be classified between quantitatively managed and optimizing for the first rubric.

2. Data acquisition, processing, and quality assurance

The overall goal of this rubric is to check if the data can be reliably captured in a way that facilitates use, preservation, and reuse based on the data management plan. This project is doing extremely well in this rubric. We give each of the sub-rubric a 5 except 2.4.3 Check data integration from other sources because all the data used by this project are collected by themselves and there is no data from other sources are used, so this sub-rubric is not applicable for this project. After all, this project received a 5 out of 5, which makes it achieve the level 5 goal, in this rubric. We want to point out that this data management plan has a very detailed guiding principle that they followed to make sure the data is valid, in addition, they also have a detailed procedure to deal with missing data. We are very impressed and inspired by their plan.

3. Data Description and Representation

The data management plan received another 5 out of 5 in this rubric, and it is doing extremely well here. The overall goal for this rubric is to check if this data management plan describes and represents data to facilitate future discovery and use. Since this project is conducted by this non-profit organization whose goal is to show everyone their research results, so the researcher team compiled a full user guide that listed everything we should know about the data they've collected, including the variables, the variants of all the variables, and even instructions of how to analyze and manipulate the data. The full user guide is about 100 pages long, they have considered every aspect of the problem that a user might encounter.

4. Data Dissemination

The overall goal for this rubric is to make sure that the research has established the policy and technical infrastructures for users to share, discover, obtain, and interact with data. This project, it wins another 5 out of 5 scores. The IEA has a completed agreement or policy for every user who wants to use their data to follow, it detailed all the possible scenarios. However, this part is not included in this particular data mana plan, but IEA has a page to publicize this information individually.

5. Repository Services and Preservation

Last but not least, this rubric is here to make sure that the research has kept the data accessible, even as hardware, software, and storage media change. IEA has published the data on their own website for us to download freely, including the previous version, they also listed another website that has IEA's authentication to download the data. However, we did locate the information about how they exactly store the data and provide the security. But we still believe that they should have this somewhere and they just didn't public it. We think they did this actually for security purposes. Even so, we deducted points from this rubric.

Conclusion

In conclusion, we think that the research did an overall good job fulfilling the expectation of the CMM model with very detailed assessment plans. Not only did it contain a very detailed assessment plan, but it also ended up with relatively fair data which helped data scientists to analyze the students of the different countries as a whole. However, at the same time, we also encourage the researchers to consider more diverse procedures for evaluation, be more transparent regarding data sharing and communicate with the participants about data ownership.

Reference

International Association for the Evaluation of Educational Achievement. International Civic and Citizenship Education Study, 2016. Inter-university Consortium for Political and Social Research [distributor], 2018-10-24. https://doi.org/10.3886/ICPSR37147.v1

Teachers College Columbia University. (2014b). Information Security Charter. Teachers College - Columbia University

Teachers College Columbia University. (2014a). Data Classification. Teachers College - Columbia University.

Qin, J., Crowston, K., Flynn, C., & Kirkland, A. (2014). A Capability Maturity Model for Research Data Management. wiki. Retrieved June 25, 2014, from http://rdm.ischool.syr.edu/xwiki/bin/view/CMM+for+RDM/1.1

Van de Vijver, Fons & Jude, Nina & Kuger, Susanne. (2018). Challenges in International Large-Scale Educational Surveys.

Appendix A – Capability Maturity Assessment Rubric Tables

	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
1. Data Management in General						
No steps have been taken to establish organizational Quantitative quality goals have been established regarding for this project, but have not taken	Stakeholder and end user needs Processes regarding stakeholder and end user needs		Stakeholder and end user needs stakeholder and end user needs	policies or senior stakeholder and end user needs	and objectives have been and end user needs and	The project follows approaches to and objectives that
1.1.1 Identify stakeholders						X
1.1.2 Develop user						X
1.1.3 Establish quantitative objectives for data						X
1.1.4 Develop communication						X
Structures or No steps have been taken to structures or budgets, staffing, or tools have been	Quantitative quality goals have been established regarding Structures or plans, training, and structures or plans, training, and		Processes regarding resources such as budgets, plans, training, and resources such		The project includes provide organizational resources such as budgets, plans, training, and resources such	as budgets,
1.2.1 Develop and implement a	X					
1.2.2 Staffing for data management						X
1.2.3 Develop collaborations						X
1.2.4 Train researchers and data management						X
1.2.5 Develop RDM tools						X
1.2.6 Establish a data management						X
No steps have been taken for Workflow management during the research process, such as managing functional requirements,	Workflow management during the research process, such as managing functional requirements,		Quantitative quality goals have been established regarding workflow during the research process, such as managing collaboration, creating		Processes regarding workflow The project follows approaches to managing the workflow during research process, such as managing functional requirements,	
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and	Focus on process
1.3.1 Manage RDM Requirements	being observed	intuitively at project level	projects and often reactive	the organization/com	controlled	improvement
1.3.2 Manage Collaborations						X
1.3.3 Create Actionable RDM						X
1.3.4 Develop Workflows and						X
No steps have been taken to measurement, analysis, and general have been recorded for	Quantitative quality goals have been established including Measurement, analysis, or measurement, analysis, or verification of the research process		The project follows approaches to establish procedures for verification of the research process		Processes regarding measurement, analysis, or verification of the research process in verification of the research process	

1.4.1 Measurement and						X
2. Data acquisition, processing and quality assurance						
Quantitative quality goals have been established regarding data quality and documentation	Data quality and documentation Processes regarding data quality and documentation are evaluated by individual team members, but no steps have been taken to	Data quality and documentation have been considered minimally addressed for this project, but have not taken wider	The project follows approaches to that have been defined for the			
2.1.1 Develop data						X
2.1.2 Develop data documentation						X
Quantitative quality goals have been established for resources, quality control procedures	Resources, structure, and training Processes regarding resources, The project provides resources, have	Resources, structure, and training No steps have been taken to structure, and training with regards quality control procedures have	with regards to file formats or structure, and training with regards			with regards to file formats or to file formats or quality
2.2.1 Develop data						X
2.2.2 Develop data quality control procedures						X
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
The workflow for collecting and documenting data has been established and	The workflow for collecting and documenting data has been No steps have been taken to	The project follows approaches to The workflow for collecting and addressed for this project, but has workflow of collecting and	Quantitative quality goals have been established regarding the of collecting and documenting data establish procedures		for the	Processes regarding the workflow of collecting documenting data has
2.3.1 Capture / Acquire data and data						X
Quantitative quality goals have been established regarding measurement, analysis, and documentation have been	Measurement, analysis, and No steps have been taken to measurement, analysis, and	The project follows approaches to Measurement, analysis, and measurement, analysis, and verification of data collection and	verification of data collection and establish procedures for verification of data collection and			Processes regarding measurement, analysis, and verification of data
2.4.1 Measurement and						X
2.4.2 Assure data quality						X
2.4.3 Check data integration from	N.A	N.A	N.A	N.A	N.A	N.A
3. Data Description and Representation						
Metadata development policies that have been established regarding metadata development are codified in	The project follows approaches to No steps have been taken to Metadata development has been that have been defined for the entire project, but have not taken wider team members, but nothing has		Quantitative quality goals have been recorded for this metadata development and are considered minimally by			Processes regarding metadata development evaluated on a establish organizational regular basis, as community needs or standards into community or
3.1.1 Develop						X
Structures or plans, training, and resources such as staffing and tools for metadata development	Structures or plans, training, and No steps have been taken to Processes regarding structures or plans, training, and		Quantitative quality goals have been recorded for this			The project follows includes resources such as staffing and plans, training, and resources such as staffing

3.2.1 Develop or adopt metadata specifications						X
3.2.2 Select and acquire tools						X
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
3.2.3 Develop strategies for generating metadata based on community practices • 3.2.4 Integrate						X
3.2.5 Arrange staffing for						X
3.2.6 Provide training for researchers and						X
3.2.7 Assess community data and metadata						X
Workflow management for Processes regarding workflow for workflow for metadata creation research process has been	Workflow management for No steps have been taken for workflow for metadata creation recorded for this project, but has	metadata creation during the The project follows approaches to metadata creation during the during the research process as	Quantitative quality goals have been established regarding research process has been managing the workflow of during the research process,			
3.3.1 Generate metadata according to						X
3.3.1.1 Document variables						X
3.3.1.2 Document files						X
3.3.1.3 Document the study						X
Quantitative quality goals have Measurement, analysis, or No steps have been taken to measurement, analysis, and compliance with metadata	Measurement, analysis, or been established including measurement, analysis, or establish procedures for measurement, analysis, or compliance with metadata	verification to ensure quality and verification to ensure quality and standards have been recorded for	Processes regarding The project follows approaches to verification to ensure quality and verification to ensure quality			
3.4.1 Measuring and Verifying						X
3.4.2 Ensure compliance to						X
3.4.3 Ensure interoperability with data and						X
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
4. Data Dissemination						
Policies for data sharing or Data sharing or confidentiality has have not taken organizational confidentiality and am	confidentiality have been recorded Processes regarding data sharing data sharing or confidentiality that individual team members, but a regular basis, as codified in policies or senior	The project follows approaches to No steps have been taken to been established regarding data wider community needs or nothing has been quantified or	Quantitative quality goals have been considered minimally by or confidentiality are evaluated on have been defined for the entire standards into	for this project, but have establish sharing or		
4.1.1 Develop data sharing policies						X

4.1.2 Develop policies for data rights and rules						X
4.1.3 Develop data confidentiality						X
Quantitative quality goals have Structures or plans, training, and structures or No steps have been taken to structures or structures or plans, training, and technologies for data sharing or structures or plans, training, and technologies for data sharing or			been established regarding Structures or plans, training, and resources for enabling plans, training, and resources for resources for enabling confidentiality have been recorded		Processes regarding The project includes resources for enabling enabling technologies for data enabling technologies for data	
4.2.1 Manage enabling technologies for access and						X
Workflow management for data for No steps have been taken for to been established regarding discovery, and citation, has been citation, data dissemination, including		Workflow management for data data dissemination, including managing the workflow of workflow for data dissemination, discovery, and citation, has been	Quantitative quality goals have dissemination, including sharing, workflow for data dissemination, recorded for this project, but has		Processes regarding workflow The project follows approaches dissemination, including sharing, sharing, discovery, and including sharing,	
4.3.1 Identify and manage data						X
4.3.2 Encourage sharing						X
4.3.3 Enable data discovery						X
4.3.4 Distribute						X
4.3.5 Ensure data citation						X
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
Quantitative quality goals have No steps have been taken to been established including to ensure accessibility measurement, analysis, or			Measurement, analysis, or Processes regarding establish procedures for The project follows approaches to measurement, analysis, and verification to ensure accessibility and security of data have been		Measurement, analysis, or measurement, analysis, or verification	
4.4.1 Measurement and						X
4.4.2 Verifying Implementation						X
5. Repository Services and Preservation						
Data preservation, curation, and Data preservation, curation, and No steps have been taken to preservation, curation, and community needs or		backups have been addressed for backups have been considered backups are evaluated on a backups that have been defined	Quantitative quality goals have this project, but have not taken establish organizational backups, and are codified in		Processes regarding data The project follows approaches to been established regarding data preservation, curation, and minimally by individual team regular basis, as codified in	wider
5.1.1 Develop data preservation		X				
5.1.2 Develop data backup		X				
5.1.3 Develop data curation policies						X
Quantitative quality goals have Resources, structure, and training resources, No steps have been taken to resources, structure, and technlogies or business models for technologies or structure, or training with			been established for resources, Resources, structure, and training with regards to enabling structure, and training, with structure, and training with regards to enabling technologies or business models for data preservation have been		Processes regarding The project provides with regards to enabling provide for resources, regards to enabling to enabling	

5.2.1 Appraise and select enabling						X
5.2.3 Develop business models						X
5.2.4 Develop backup procedures and		X				
	Level 0	Level 1: Initial	Level 2: Managed	Level 3: Defined	Level 4:	Level 5:
Process / Practice	This process or practice is not	Data are managed	DM process is characterized for	DM is characterized for	DM is measured and controlled	Focus on process improvement
The workflow of data preservation, including storage, security, version workflow No steps have been taken to preservation, workflow of data preservation, security, version addressed for this project, but has security, version control, workflow of data preservation, control, and migration, has been not taken wider community needs control		The project follows approaches to The workflow of data preservation, including storage, security, version	Quantitative quality goals have been established regarding the of data preservation, including establish procedures for the		Processes regarding the the workflow of data including storage, storage,	
5.3.1 Store data		X				
5.3.2 Provide data		X				
5.3.3 Control changes to data						X
5.3.4 Backup data		X				
5.3.5 Curate data						X
5.3.6 Perform data		X				
Quantitative quality goals have been established regarding No steps have been taken to Measurement, analysis, and measurement, analysis, and backups have been recorded for		The project follows approaches to Measurement, analysis, and measurement, analysis, and verification of data storage and			Processes regarding measurement, analysis, and verification of data storage and verification of data storage	
5.4.1 Measurement and		X				
5.4.2 Validate data		X				
5.4.3 Validate backups		X				