Measuring Girls' Agency in East Africa—Co-Creating Contextually Specific Tools for Evaluation: Lessons from the AMPLIFY Girls Collective

By Aubryn Allyn Sidle (Cornell University), Duncan Ahigika (Elohim Development Association), Wendo Aszed (Dandelion Africa), Shanu Bashuna (Chalbi Scholars), Samantha Belle (AMPLIFY Girls), Dativah Bideri (Komera), Elizabeth Bohart (AMPLIFY Girls), Margaret Butler (AMPLIFY Girls), Kansiime Honest Fortunate (Girls to Lead Africa), Zachary Fowler (WISER), Linda Kamau (AkiraChix), Jess Littman (AfricAid), Valentine Makamuyenzi (Streets Ahead Children's Center Association), Estahappy Mariki (The Girls' Foundation of Tanzania), Anande Mirisho (Jifundishe), Phionah Musumba (Malkia Foundation), Maria Omare (The Action Foundation Kenya), David Omondi (Riley Orton Foundation), Brenda Oulo (AMPLIFY Girls), Dorcas Oyugi (WISER), Elly Sarakikya (Sega Girls Academy), Kimberly Wolf (GirlUp Initiative Uganda).

Introduction

Integrating the voices and perspectives of community-driven organizations into global policy discussions is a growing priority in international development. While some strides have been made to shift funding to grassroots organizations, much work remains to fully recognize the power, wisdom, and programmatic outcomes of local organizations in global conversations about best practice. This recognition is imperative, given that community-driven organizations are best positioned to understand the complexity of local context and to holistically address the challenges facing communities. Thus, there is no substitute for their expertise.

We argue that the neglect of local voices is driven, by both the technical processes that define and measure success, combined with a prioritization at the global level of scale over quality of impact. Most international development initiatives rely heavily on quantitative indicators to measure programmatic impact. These measures are usually developed without the input of local actors, whose technical capacities to implement such indicators are often limited, and whose sizes are usually too small to warrant a seat at the table. The scarcity of grassroots organizations' involvement in these processes highlights a troubling reality; grassroots organizations have been shut out of the discussions that measure their own effectiveness and success. By doing this, global actors have denied community-driven organizations a seat at the table where success is defined, best practices are established, and ultimately, where investment strategies are created.

Our approach offers an alternative to traditional top-down measurement processes. Our work develops a tool for quantifying impact, but does so by prioritizing the perspectives and expertise of local actors. This research is the result of a collaborative partnership between AMPLIFY—a collective of East African community-driven organizations focused on improving the lives of adolescent girls—and Aubryn Allyn Sidle, a researcher and PhD candidate from Cornell University. Our research partnership had three over-arching goals: to collaboratively define success for girls, construct a contextually responsive measure of that success, and document collective impact.

Utilizing Group Concept Mapping (Trochim and Kane 2005), AMPLIFY members identified girls' Agency as the key outcome of interest for their communities and therefore the focus of collective efforts to document impact. This paper provides a detailed description of our process

and methodology for developing a quantitative measure of Agency. Our research is informed by three basic assumptions: First, *Agency* can be taught and is being taught well to adolescent girls by many grassroots organizations. Second, agency is a complex and contextualized capacity—influenced by the opportunities and constraints of each community. Similar to the notion that the value of vocational skills to youth employment is dependent on the structure of local and regional economies, we believe *Agency* is also not a monolithic construct, but one that is shaped by the realities of place. This brings us to our final point, that the organizations best positioned to teach Agency are community-driven, precisely because they best understand the specific challenges of local context that might constrain or enable agentic development.

Context: AMPLIFY—A Community Driven Solution to Adolescent Girls' Success

AMPLIFY was founded in 2017 by *Komera* & the *Antelope Foundation*, to demonstrate the power and impact of community-driven organizations in serving adolescent girls. The idea of AMPLIFY was further developed and brought to fruition in partnership with 18 organizations across Eastern Africa. Focused on scaling local solutions for adolescent girls through intentional collaboration, AMPLIFY is a collective of these 18 organizations who are located across Tanzania, Rwanda, Uganda and Kenya. AMPLIFY partners independently design and implement a diverse set of programs to deliver a holistic model of care for young women that focuses on transformational change and the documentation of collective impact.

AMPLIFY partners are located in both urban and rural communities, and utilize multiple types of interventions situated in a variety of sectors including; education, health, economic empowerment, and community development. The scope and diversity of partner organizations contribute greatly to the strength of AMPLIFY's work as a collective. (For a detailed description of AMPLIFY members' programs, please see appendix 1). AMPLIFY engages the Executive Director from each partner organization in a general assembly (GA) and is governed by an elected Board derived from the membership. The GA utilizes a consensus-based decision-making model to move the work of the collective forward and determine organizational strategies.

Innovating to Define Success for Girls

During its inaugural year, AMPLIFY identified the need for a monitoring and evaluation strategy as its key priority for action. Establishing a set of shared metrics across the collective was achieved in 2018 through **Group Concept Mapping—**a participatory mixed-methods tool that supports the visualization and synthesis of multiple perspectives (Trochim & Kane 2005). Group Concept Mapping began by asking AMPLIFY members a simple question: **What does success look for girls in your community**?

> *Girls increase their self-confidence. Girls are proud of who they are. Girls can set and achieve goals. Girls can see that they can achieve what they want if they put in the work. Girls have the skills to make informed decisions. Girls are able to decide when they get married. Girls are able to create positive change in their environment. Girls influence policy and implementation.* (AMPLIFY members, Nairobi Meeting, October 2018).

A clear theme emerged - success for girls is characterized by the ability to effect change in their own lives and world. This capacity was identified by AMPLIFY members as *girls' Agency* and was unanimously ranked as the most important outcome from their collective work—followed by (in order of priority) formal schooling, economic empowerment and girls' health and freedom.

Psychologist Albert Bandura defines agency as the ability to "influence one's functioning and the course of events through one's actions."[1] Whereas Agency has been identified as a skill of interest for education (Kwauk & Braga 2017; Murphy-Graham & Lloyd 2016) and gender empowerment more broadly (Kabeer 1999), a rigorously tested psychometric measure for assessing agency doesn't yet exist. Further there is no consensus on whether or not agency can be 'taught'—particularly to older adolescents (Kautz et al. 2014)—or whether it is largely pre-determined by the structure of the environment in which girls' live (Kwauk & Braga 2017).

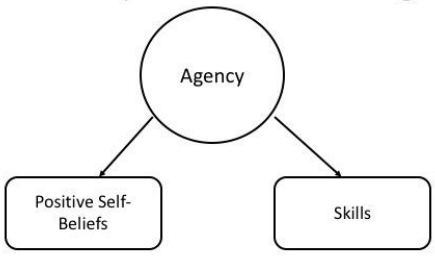
It follows then, that *accurately* measuring something as complex, and contextually dependent as *Agency,* is a daunting challenge for large international organizations operating from a ten-thousand-foot multi-country lens. While community-driven grassroots organizations are better positioned with the local expertise to take up this challenge, they rarely have the technical capacity or resources to do so. Thus, our efforts began with providing technical training to grassroots organizations, and technical support throughout the course of our research. We utilized a participatory mixed-methods approach to develop and validate a psychometric measure of girls' *Agency.*

Utilizing Participatory Practice to Define and Situate Agency in Context

Agency is a deeply contextualized construct—dependent upon how individuals view themselves, their communities, and the specific skills necessary for taking effective action given the opportunities and constraints of each place. As such, *Agency* is sensitive to gender, culture, context and other identity-related markers. Because these experiences are not easily quantifiable, it is necessary to allow qualitative input to drive quantitative measurement processes in order to accurately capture capacities which vary based on people's perceptions, social realities, and lived experiences (Hillenbrand 2015, Narayan 2005). This underscores the need for participatory approaches that could situate the measurement of soft skills like *Agency,* in local socio-cultural realities.

Of particular importance to our study is the question of accuracy and relevance, and the documented ability of participatory practice in evaluation to both improve accuracy of measurement, and contextualize results (Newton, et al., 1999; Parrado, McQuiston & Flippen, 2005). This is a particularly salient need for the AMPLIFY collective whose diversity of programmatic approaches for producing girls' *Agency* required a contextually flexible definition. In her

Figure 1: Conceptual Definition of Agency



[1] https://albertbandura.com/albert-bandura-agency.html

paper "Action on Agency" (2019), Sidle offers such a definition, by synthesizing the literature on agency from the human development, psychology, and education fields and adapting it for use by practitioners. Sidle identifies *Agency* as "the capacity of individuals to define aspirational goals and coordinate the [necessary] knowledge, skills, attitudes and resources…to take action to achieve stated goals. Thus, Agentic capacity is made up of both positive self-beliefs and concrete skills" (4-5). See figure 1 (left).

By not identifying the specific self-beliefs and/or skills that *must* make up *Agency*, this definition, allows for Agency to be adaptable to context—where each organization can identify the skills and beliefs that matter for girls in their local communities. For AMPLIFY, the outcome area of "Girls' Agency" was derived from 36 individual responses from member organizations. Thematic coding of these responses clearly reflected the two broad categories of beliefs and skills. Thus AMPLIFY members' understanding of agency mirrored Sidle's conceptual definition, indicating that this definition of agency (as skills plus positive self-beliefs) was a good starting point for measurement. To identify the specific skills and beliefs that would be included in AMPLIFY's Agency measure, we conducted a series of conversations and surveys with AMPLIFY members as described in the next section.

Methodology: A Participatory Mixed Methods Approach

Each stage of research and the creation of the Sidle-AMPLIFY *Agency* tool was initiated by AMPLIFY members, driven by member input, and, ultimately, limited only by the collective's decision to discuss further or proceed. During these discussions, researcher facilitation assisted the collective in collaborative decision-making but did not influence the decision itself. Researchers sought to establish a "space of control" where stakeholders were recognized as experts and could affect change in the processes by which their expertise was heard and recorded (Nelson 1995, Giddens 1984, Newton 2019). Participatory methods were combined and interwoven with more traditional investigation of statistical validity described below.

Burton and Mazerolle (2011) lay out a straightforward methodology for validation of psychometric measures. In their paper, "Survey Instrument Validity Part 1" they identify four types of validity that are important for establishing credibility of a new measure: content validity, face validity, criterion validity and construct validity (ibid, 28). We established both content validity (ensuring that the measure has the right 'content' or is representative of the totality of thing being studied) and face validity (the readability/understandability and relevance of the measures to the content), through a combination of literature review and participatory processes with AMPLIFY members and their constituents.

While establishing content and face validity are often qualitative processes (ibid), establishing the construct validity of a measure–in other words confirming that the measure does indeed measure what was intended—is typically done through statistical analysis of hypothesized relationships between constructs (Cronbach & Meehl 1955). We analysed the construct validity of our measure by utilizing confirmatory factor analysis (CFA) as suggested by Burton and Mazerolle (2011) to assess the fit of our hypothesized conceptual definition of Agency. To strengthen our evaluation of construct validity through participation, we involved AMPLIFY members in building the hypothesis of the underlying relationship between constructs and thus refining Sidle's conceptual definition.

Lastly, we wanted to note, that limited work was done to assess criterion validity which seeks to establish the extent to which a new measure compares to an existing, similar, measure that is seen as the 'gold standard' (Burton & Mazerole, 29). Since there is no pre-existing measure of Agency, or related construct, criterion validity was difficult to assess. Although we acknowledge that future analysis which seeks to understand the predictive power of our measure against something tangentially related such as 'locus of control,' would be a valuable undertaking. The complete process of survey development and validation is detailed in the following paragraphs.

Content and Face Validity: The Process of Survey Development

The content validity of the measure was established by asking AMPLIFY members, as experts, to identify the specific self-beliefs and skills that are important for Agency in their contexts. Using a written survey, members were asked to indicate skills and self-beliefs that they believed were important for girls' success AND that were explicitly taught by their programs. A final list of 13 skills and 3 self-beliefs was compiled from this survey. Researchers reviewed the literature for existing psychometric measures of each and drew from pre-existing scales to produce Draft 1 of the survey (described below).

The Rosenberg Self-Esteem Scale (Rosenberg 1965), general self-efficacy scale (Shwarzer & Jerusalem 1995) and empowerment scale as conceived by Torre (1986) were used to measure the self-beliefs portion of the survey. Skills drew from a variety of pre-existing measures, some of which were kept intact, and some of which were reviewed for individual questions of merit. Scales for measuring goal setting, interpersonal communication and decision making as developed by Mincemoyer & Perkins for evaluation of life skills programs in the US (2005) were used as-is. Questions were adapted from CARE's Gender Equitable-Men's Scale (2015) for measuring gender attitudes and gender-based violence awareness. Questions on perseverance were drawn from Duckworth & Quinn's Grit scale (2009) measuring 'perseverance of effort,' while the self-awareness measure was drawn from the self-consciousness scale developed by Scheier & Carver (2013). Several skills did not have pre-existing psychometric measures or measures that matched the way AMPLIFY members had defined them. In these cases, researchers drew from members' existing survey and survey questions. Specifically, questions measuring 'volunteerism' (both mentoring and community engagement), gender rights literacy, leadership, and public speaking were drawn from questionnaires developed by *AfricAid* in Tanzania,[2] one of AMPLIFY's member organizations.

In April 2019, at the bi-annual meeting of AMPLIFY in Arusha Tanzania, members reviewed Draft 1 of the survey. During workshops and focus group discussions, members were asked to define the skills/self-beliefs on the survey from the perspective of their local community context and identify whether or not the selected survey questions matched these definitions—assessing each item for its *face validity* (did the item match the definition of the construct?). Several of the sub-scales were revised heavily by this process, as it was determined that members' definitions of the skills did not match the way the survey defined them. Among those facing heavy revision were interpersonal communication, self-awareness, gender attitudes, financial literacy/entrepreneurship and civic engagement. Interestingly, members felt the self-

[2] *AfricAid* is locally known in Tanzania as '*GLAMI.*'

belief questions (measuring self-esteem, self-efficacy, and empowerment) were accurate, and these questions remained un-revised.

Member revisions resulted in an overall reduction of the skills from 13 to 11. Financial literacy, entrepreneurship, and civic engagement were dropped from the survey after partners agreed that these were important life skills but were not necessarily essential for Agency. Gender attitudes, gender rights, and gender-based violence awareness were combined to make one sub-scale called gender positive attitudes. After much discussion, members added 'problem solving' and 'conflict resolution' to the survey based on the perspective that these skills were particularly relevant for Rwanda-based organizations, who viewed the ability to resolve inter-personal conflict as critical to success in their communities. Added questions were developed by AMPLIFY members themselves at the meeting and were not drawn from pre-existing measures. The Tanzania revisions resulted in Draft 2 of the survey.

Draft 1 and Draft 2 (as described above) of the survey were piloted in English with respondents in Tanzania and Uganda, respectively, for a total of 334 respondents ranging in age from 12-35. Statistical analysis of early pilot data was used to further revise the survey and reduce the length. The data reduction procedure was conducted as follows: first, the distributions of each individual question were considered and questions with little or no variability were dropped—indicating that girls either did not understand the question or would likely answer the same way. Exploratory factor analysis was conducted on each of the sub-scales to determine the number of latent factors for each, and for those scales with one latent factor, items with factor loadings lower than .3 were dropped. *The final Sidle-AMPLIFY Agency survey is a 78-item instrument designed through a collaborative process to ensure content, face and construct validity and pilot testing of an initial version with revisions*. See Table 1 (below) for the individual skills and self-beliefs included in Draft 1 of the survey compared to the final.

Table 1: Overview of Skills and Beliefs in Draft 1 Compared to Final Draft of Sidle-AMPLIFY Agency Survey

Survey Draft 1	**Final Sidle-AMPLIFY Agency Survey**
Beliefs	Beliefs
Self-Esteem	Self-Esteem
Self-Efficacy	Self-Efficacy
Empowerment	Empowerment
Skills	Skills
Volunteerism (mentoring & community engagement)	Volunteerism (mentoring & community engagement)
Leadership	Leadership
Interpersonal Communication	Interpersonal Communication
Decision Making	Decision Making
Public Speaking	Public Speaking
Goal Setting	Goal Setting
Self-Awareness	Self-Awareness
Perseverance	Perseverance
Attitudes towards gender	Positive Gender Attitudes
Gender Rights	

Ability to Report/Identify GBV	
Financial Literacy	Conflict Resolution
Entrepreneurship	Problem Solving
Civic Engagement	

Establishing Construct Validity:

Next, the construct validity and internal reliability of the agency instrument were statistically assessed. Internal consistency was determined utilizing exploratory factor analysis (EFA) and associated measures of internal consistency or alpha coefficients. In other words, we assessed each set of questions measuring a specific skill or self-belief on the survey for the degree to which those questions could be statistically viewed as a reliable measure (this is expressed by the Cronbach's alpha statistic).

Secondly, we hypothesized a conceptual model for how these skills and self-beliefs fit together to constitute agency. Next, we tested the conceptual model statistically by using a process called Structural Equation Model (SEM) as a form of Confirmatory Factor Analysis (CFA), to see if our hypothesized conceptual model was reflected by the data. This process effectively establishes the construct validity of the instrument (Burton & Mazerolle 2011). Our work follows Babyak et al (1993) and their validation of the Snyder Hope Scale. The structural equation model was developed based on a conceptual framework constructed in collaboration with AMPLIFY members and exploratory factor analysis as described below.

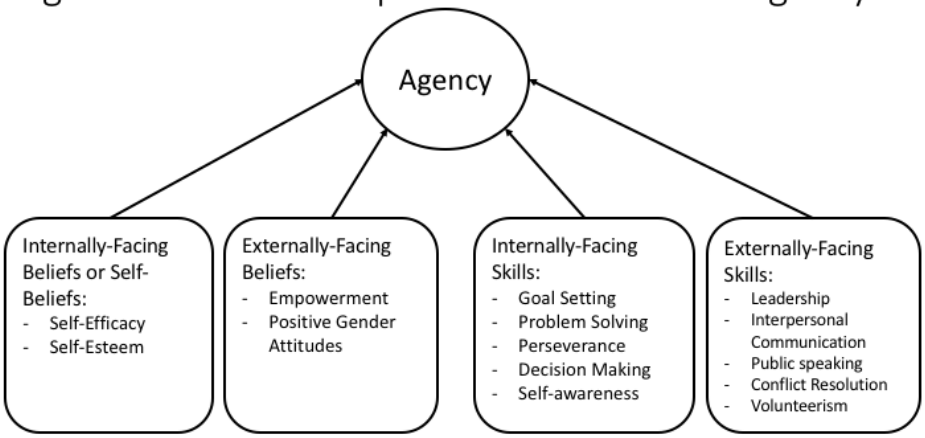
To build our conceptual model for CFA, we identified the combinations of skills and self-belief questions that we believed were part of the same underlying construct. We began with the base model developed by Sidle (2019, see figure 1), which informed the initial structure of the *Agency* survey. This simple model was further refined by AMPLIFY members in workshop sessions where groups of 3-4 members, were asked to combine the skills and beliefs included in the survey into conceptual categories. Each groups' final categories were presented to the General Assembly for discussion and revision. The researchers took the four final configurations of skills and beliefs and synthesized them into one version based on themes, majority opinion, and instruction from literature.

This synthesis was then statistically tested utilizing exploratory factor analysis (EFA) as described in the previous section. Skill and belief groupings were assessed for whether or not they had one latent factor (as demonstrated by eigen values above 1 explaining at least 70% or more of the variability). Items (individual questions) with factor loadings lower than .3 were dropped. Internal consistency was calculated for each group category utilizing the Cronbach's alpha statistic. We examined the coefficient alpha index adopting the general convention in research as prescribed by Nunnally and Bernstein (1994), which states one should strive for reliability values of 0.70 or higher.

This collaborative process resulted in the final Conceptual Framework for the *Agency* tool which further divides both skills and beliefs into two additional sub- categories—those that are internally facing and those that are externally facing (see Figure 2). First order latent variables

were created by combining the four categories of skills and beliefs found to have a single underlying factor into first order latent variables predicted from the factor loadings. Our Confirmatory Factor Analysis (CFA) hypothesized a simple structural model, using the first-order variables to test the second order latent variable, Agency (figure 2).

Figure 2: Final Conceptual Framework for Agency



To assess the fit of the Final Conceptual Model (figure 2) the following fit indices were used; the chi-square with degrees of freedom, root-mean-square error of approximation (RMSEA), and Comparative Fit Index (CFI). Chi-square tests in this context, evaluate whether or not the model is statistically different from the data. Model fit is ideal when there is no difference between data and the model. According to Schermelleh-Engel et al. (2003) a value of RMSEA of about 0.05 represents a close fit of the model, although values up to 0.08 may be acceptable. The RMSEA represents the amount of error in the model. We present both conservative and liberal cut-offs (RMSEA $\leq$ 0.06-conservative, or $\leq$ 0.10-liberal) for an acceptable fit because fit indices may be affected by factors such as sample size and data distributions (ibid). Similarly, and also following Schermelleh-Engel et al, we evaluate the CFI based on a both liberal and conservative cut-offs. A model that adequately fits the data will have CFI $\geq$ 0.95 (conservative) or $\geq$ 0.90 (liberal).

Table 2: Data Demographics	
Total by Country	
Kenya	690
Tanzania	394
Rwanda	269
Uganda	124
Total by Urban/Rural	
Urban	239
Rural	1,238
Total by Baseline/End line	
Baseline	788
End line	688

Age	
Age range	10-38
Mean Age	17
Total Observations	1,477

The final model was adopted following satisfactory fit statistics (as described above). A collective *Agency* score was predicted and tested for normal distribution. Validity was further investigated by testing the hypothesis that individuals joining programs at baseline have lower scores of *Agency*, compared to those completing programs. Stata 16 was used to specify, estimate, and evaluate the models.

Data:

Analysis for CFA was conducted on data gathered from 15 of the 18 AMPLIFY members for a total of 1,477 observations. The data was cross-sectional representing either students at baseline (just entering AMPLIFY members' programs) or at end line (in their final few months/weeks of members' programs). A simple description of the data demographics can be found in Tables 2 & 3. Overall, ages ranged from 10 to 38 with a mean age of 17. While data was collected from all four AMPLIFY countries, almost half of the respondents were from Kenya (47%). The majority of respondents were from rural communities (1,238), while 239 respondents were from Urban or peri-urban areas. Just over half the respondents represented students at baseline (the balance at end line). Six hundred twenty-four respondents or 45% came from families where parents had a primary school education or less, 410 or 28% reported having parents who had attended or completed secondary school, and 18% (266 respondents) had parents who had completed some higher education (see table 3). Another 10% of respondents had no recorded data for their parents' highest level of education.

Table 3: Parental Education	
Parents' Highest Level of Education	Number
uneducated	169
some primary school	34
primary school	421
some secondary	28
Secondary school	365
some A-level	1
A-level	21
college (diploma)	89
university	98
masters	4
PhD	3
Total	**1,234**

Results:

Alpha coefficients were examined for each individual belief or skill sub-scale on the *Agency* instrument, in addition to the combined factors described above. A single factor was found for each category identified in figure 2 and examination of internal consistency for each of the four domains was found to be more than satisfactory.

Results of exploratory factor analysis showed each of the 4 belief sub-scales (self-esteem, self-efficacy, gender beliefs and empowerment) as well as 2 of the skill sub-scales (leadership and self-awareness), showed one latent factor with an eigen-value of 1.0 or higher. Cronbach's Alpha for each sub-scale is presented in Table 4. We found that internal reliability of individual sub-scales ranged from .34 on the low-end to .75 (self-efficacy) on the high-end with six out of fourteen sub-scales demonstrating internal reliability of .6 or higher (self-efficacy, self-esteem, empowerment, gender beliefs, leadership and self-awareness). These same six also showed a single latent factor. An additional 2 sub-scales (public speaking and volunteerism) had alphas of .56 with rest of the sub-scales presenting low reliability below .5 (decision making, perseverance, problem solving, goal setting, conflict resolution, and interpersonal communication). All sub-

scales with internal reliability scores below .5 showed no evidence of any latent factors with eigen-values below the 1.0 cut-off.

We interpret these findings to mean any individual scale or combined skill grouping (self-beliefs, environmental beliefs, etc) that have an alpha of .7 or higher demonstrate satisfactory reliability. Those with alphas above .6 but below .7 indicate good, but not satisfactory reliability. In both cases, we think these scales do represent a single measure which warrant further investigation. We do not consider those below .5 to be reliable measures.

Table 4: Average Inter-item Correlation and Reliability of Sub-Scales and Combined Sub-Scales

Sub-scale	# of items	Average interitem correlation	Standardized alpha score
Self-Esteem	7	0.204	0.62
Self-Efficacy	8	0.279	0.75
Self-Beliefs	**15**	**0.204**	**0.79**
Empowerment	9	0.163	0.61
Gender Beliefs	11	0.15	0.63
Externally Facing Beliefs (Environmental Beliefs)	**20**	**0.142**	**0.73**
Goal Setting	3	0.161	0.34
Perseverance	4	0.162	0.45
Decision Making	4	0.225	0.49
Problem Solving	3	0.224	0.49
Internally Facing Skills (Critical Thinking/Self-Governance)	**14**	**0.175**	**0.74**
Leadership	5	0.278	0.67
Interpersonal Communication	5	0.117	0.42
public speaking	3	0.283	0.56
Conflict Resolution	5	0.132	0.44
Volunteerism (community Engagement & Mentoring)	4	0.258	0.56
Self-Awareness	6	0.236	0.63
Externally-Facing Skills (Leadership)	**28**	**0.18**	**0.85**

Analysis of the four combined domains of skills and beliefs from our conceptual framework is also shown in Table 4. Each of these skill and belief groupings showed only one latent factor

per grouping with group internal reliability scores ranging from .73 to .85 for each set. When each group was combined into a single latent factor variable predicted from factor loadings of each item, these combined variables showed even higher degree of reliability ranging from .828 to .898 (see Table 5).

Table 5: Average Inter-Item Correlation and Reliability of Combined Latent Variables		
Latent Factor	**Average interitem correlation**	**Standardized alpha**
Self-Beliefs	0.6468	0.846
Externally-Facing Beliefs (Environmental Beliefs)	0.7456	0.898
Internally-Facing Skills (Critical Thinking/Self-Governance)	0.6165	0.828
Externally-Facing Skills (Leadership)	0.6407	0.843
Test scale	0.6624	0.887

Confirmatory Factory Analysis:

We tested our questionnaire as a comprehensive measure of *Agency* which includes four latent factor variables empirically derived as described above. Our hierarchical structural equation model posits that all items on the agency questionnaire load onto the four distinct uncorrelated first order latent variables of self-beliefs, beliefs about environment, internally facing and externally facing skills. We tested the final conceptual framework depicted in Figure 2 as a SEM with four observed variables representing the predicted scores of the first order latent variables and one latent score, *Agency*. Satisfactory fit of our conceptual framework, using CFI and RMSEA statistics, as illustrated in Table 6, was assessed. Chi square and TLI statistics are also illustrated in Table 6. There were no suggested modification indices of the final tested model. The full display of the structural equation model can be viewed in appendix 3.

Table 6: Model Fit Statistics of CFA

Assessing model fit statistics				
	Chi square statistic	**RMSEA** ≤ 0.06 (conservative) or ≤ 0.10 (liberal)	**CFI** ≥ 0.95 (conservative) or ≥ 0.90 (liberal)	**TLI**
Self- beliefs	0.000	0.06	0.857	0.833
Environmental beliefs	0	0.066	0.692	0.656
Internally facing skills	0.000	0.038	0.907	0.896
Leadership/externally facing skills	0.000	0.046	0.862	0.848
Agency	0.167	0.028	0.999	0.998

Our RMSEA fit statistics were found to be satisfactory. CFI statistics were found to be within range for internally facing skills (.907), just outside the liberal cut-off for self-beliefs (.857) and leadership skills (.862) but well below the liberal cut-off for environmental beliefs (.692). However, TLI and CFI scores were highly satisfactory for the overall model at .99. Examination of the distribution of predicted agency scores is normal. In addition, analysis of baseline versus end-line *Agency* scores showed an average increase in results over time—further indicating the validity of the measure.

Discussion & Limitations: Synthesizing Measurement of Life Skills

We'd like to highlight several key contributions of these findings. First is the relationship between Agency and structure. Much has been written about the environmental structure that constrain the Agencies of women and girls (Kwauk & Braga 2017, Maslak 2008, Kabeer 1999). Thinkers from the field of psychology posited that development and the execution of personal Agency is directly related to the structure of environment (Lerner et al. 2005, Bandura 1999). This makes sense, given that most define *Agency* as including the ability to affect change in your environment and circumstances (Little et al 2006, Bandura 2009). Whereas environmental constraints are mostly external to the individual, beliefs about its malleability are arguably central to *Agency* as a capacity—meaning that a clear indication of a person's *Agency* is the extent to which they *believe* that the circumstances they face are surmountable. Importantly, our modelling of *Agency* represents these environmental constraints in the latent factor of 'externally facing beliefs,' which include measures for empowerment and gender positive attitudes.

AMPLIFY members defined empowerment as 'belief in your ability to change your circumstances'—which is directly related to the idea of externally facing beliefs. The survey questions grouped as 'gender attitudes' are also relevant to the environmental constraints of girls' *Agency* since gender norms form the basis of socio-cultural structure in which girls' lives exist and are ordered. Thus, gender attitudes combined with empowerment in the Sidle-AMPLIFY instrument effectively measure a construct we call 'positive environmental beliefs' or, in other words, the extent to which an individual believes her environment is changeable is ultimately related to how rigid she believes that structure is constituted.

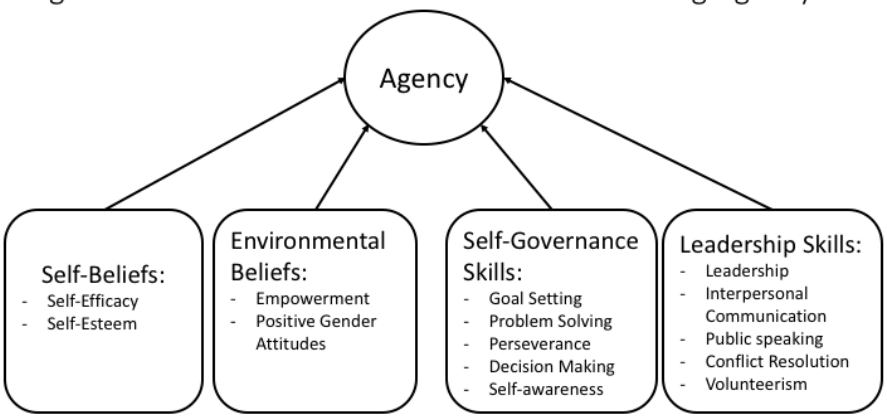
We note, however, fit statistics for Environmental Beliefs latent factor were not ideal. The CFI statistic for Environmental beliefs was .692 below the liberal cut off of .9, although RMSEA scores were well within acceptable range. Although, conceptually, we maintain environmental beliefs is a good fit for our measure of *Agency,* (and the fit statistics for Agency as a whole support this claim), the CFI statistic for environmental beliefs suggest that there is some instability with this particular construct which could be strengthened and thus warrants further investigation. We hypothesize that such instability with 'positive environmental beliefs' is likely a result of the changing nature of gender norms across communities.

Second, the combination of internally and externally facing skills are also valuable categories, synthesizing by our measure many of the individual skills taught in program practice, into two evaluative constructs. When considered at face-value, externally-facing skills as a whole represent the skills required to make change in the social environment, including: interpersonal communication, conflict resolution, public speaking and volunteerism. Volunteerism as defined by the survey measures two core capacities: mentoring and community engagement.

Broadly constituted, these skills relate to intra and interpersonal communication, propensity for, and interest in, leadership positions and mentorship, and abilities related to influencing and moving groups or teams. We argue these basically correspond to "leadership," identifying this latent factor as 'leadership skills.'

Internally-facing skills are less easily labelled. There is some evidence from education literature to suggest the skills encapsulated by this construct are all critical thinking skills. Stella Cottrell (2017) identifies perseverance, and the ability to assess and weigh evidence (more or less corresponding to decision making and problem solving) as two of many distinct skills necessary for critical thinking. Our measure of internally-facing skills includes decision making, problem solving, perseverance, and goal setting as the sub-scales making up the latent factor, but is not as broad as Cottrell's definition of critical thinking. We argue that internally-facing skills correspond more directly, perhaps, to something called self-governance which was the title given to this category of skills by AMPLIFY members during conceptual framework discussions in August 2019. Based on this input and reflecting our commitment to privilege partners' expertise, we revise our final conceptual framework to a simple model that includes the titles identified in Figure 3.

Figure 3: Sidle-AMPLIFY Framework for Measuring Agency



Third, while many of the individual scales on skills and beliefs did not demonstrate satisfactory reliability on their own, one clear contribution of this paper is the combined latent factor measures, created by our Final Conceptual Framework (figure 3). Whereas low reliability could be a result of inadequate questions, or measures with low face of content validity, it could also be a result of the fact that individual skills and beliefs on their own are too inter-related to be accurately measured with this kind of an instrument. When taken at face-value many of the individual skills and psychometric constructs are deeply inter-related. It is logical to imagine, for example, that a person's self-esteem would be directly related to their effectiveness as a public speaker, or as a decision maker. Our Conceptual Framework offers guidance to practitioners for how to assess these skills, while taking into account their inter-relatedness. In practical terms, this means that practitioners can assess public speaking skills, for example, as an indicator of leadership, or goal setting ability as an indicator of self-governance.

Most importantly, however, as debate has ensued in girls' life skills programming about which skills are most important and how might they be measured, we hope this tool will be an important resource. Our analysis offers some guidance on which skills should be the focus of

programming for producing Agency. Our work indicates that individual skills are less important than the overarching capacities the skills represent, e.g. positive self-belief and leadership. From a point of practice, this would mean that it may be less important which particular skills are addressed in programming, as long as the overarching capacities of Leadership, Self-Governance, Self-Belief, and Environmental Beliefs are targeted more broadly. Similarly, the individual skills making up the latent factors can be viewed as a starting point for thinking about curricula and programming and what might produce 'Leadership' and 'Self-Governance.'

In addition to previous discussions about environmental beliefs, several other important limitations should be noted from our analysis. First and foremost, whereas the structural equation SEM model takes into account the inter-relatedness of individual skills and beliefs as represented by our four latent factor categories, the SEM currently posits that Self-Beliefs, Environmental Beliefs, Leadership Skills and Self-Governance Skills are independent from one another. For the reasons already mentioned, that is unlikely to be the case. Follow up analysis will include further exploration of the relationship between these latent factors and whether or not hypothesized relationships are supported by the data. The potential relationships between self-beliefs and skills is of particular interest. Relationships in either direction between these constructs would be instructive for practice. For example, if heightened self-belief was shown to contribute to heightened skill in either area, it would imply organizations seeking to increase girls' agency should focus on psycho-social support or other programs targeting increased self-esteem and self-efficacy as the first priority.

Secondly, our current sample was not perfectly representative of all AMPLIFY members or all countries. Observations from Kenya made up 47% of the total, compared to Uganda (at the low-end) representing just 8% of the overall data. Tanzania and Rwanda made up 27% and 18% of the data respectively. The sample was also skewed towards participants living in rural areas (84%) compared to 16% in urban or peri-urban settings. While the urban-rural divide is largely representative of the AMPLIFY demographic as a whole, the country distribution did under-represent Uganda, Rwanda and Tanzania. [3] At a minimum, this suggests that validation results may not be as strong in Uganda and, possibly, Rwanda, where sample sizes were particularly low. We hope to include in future analysis a country by country validation.

Related to the issue of country and under-scoring the need for country by country validation, is the question of language. The *Agency* survey was translated into three languages from the original English (Swahili, Kinyarwanda and Luganda) and administered in three (English, Kinyarwanda and Swahili). The majority of Tanzanian respondents took the survey in Swahili and the majority of Rwandan respondents took the survey in Kinyarwanda. Some Kenyan respondents took the survey in Swahili, but the majority took it in English. All Uganda respondents took the survey in English. In addition to potential discrepancies between understanding of the Swahili in Kenya versus Tanzania, or understanding of the English between Kenya and Uganda, it is also possible that younger students who took the survey in English may not have fully understood all questions—since many students in these countries do not typically achieve English fluency until later in secondary school. The language version offered to participants, was decided by organizations themselves, after consultation with the researchers.

[3] At the time of writing, additional data was still being collected in these three countries and will be incorporated into future analysis.

Finally, we note the entire survey is a self-reported measure. Since many of these skills and beliefs are difficult or impossible (in the case of beliefs) to objectively observe, we felt self-reported measure was warranted. We acknowledge this measure of Agency could be strengthened through the triangulation of objectively observed measures by teachers or program staff of observable skills. However, we would argue for small and grassroots organizations, reliable objective measures are beyond the capacity of most organizations to develop and/or execute. They are similarly difficult to operate at scale and so may not provide a practical solution for the AMPLIFY constituency.

Conclusion

As part of the participatory practice of creating the Sidle-AMPLIFY Agency tool, partners' reflections have served as a vital point of feedback during this process. Positive experiences largely focused on feeling as if historically devalued voices were given space and power to influence decisions and on believing that the process was appropriately equitable for all partners, regardless of location or previous experience with survey tool development. This has not only created buy-in for widespread uptake of the survey, despite it being a heavy lift for many low-capacity partners, but overall satisfaction with the process and final product, which also indicates the accuracy of the measure.

Overall, the participatory model was well-received among partners with many of the participants noting that the method offered a welcome alternative to long-held beliefs on how to handle multinational collectives. One participant offered:

> *Often, for the sake of efficiency, a facilitator will summarize thoughts and feelings in the room too quickly, resulting in important conversations being cut off or in making some participants feel they did not have the chance to voice their opinion. In other spaces, the leadership will stop the conversation before a decision is made, saving that decision for themselves, choosing to do that between sessions rather than during one. The way in which everyone was held as a valuable and equitable contributor in the space made this process feel different – like the collective was made up of experts directing a process based on their own experiences.*

Moving forward, the process by which this survey tool was developed will serve as the basis for other collaborative efforts between AMPLIFY partners and ongoing research. We believe that "women's organizations and social movements in particular have an important role to play in creating the conditions for change" and that the process of generating Agency for women and girls "is dependent on collective solidarity" embodied through not only our interventions, but also through our evaluations (Kabeer 1999). With these thoughts in mind, we consider participatory evaluation to be instrumental in terms of accurately measuring outcomes at the local level, and ensuring that the impact of local solutions are appropriately represented in the menu of options considered in the global policy agenda.

Specifically, we argue that participatory processes will result in better ability to accurately assess progress on many of the Sustainable Development Goals (SDGs). In the education space for example, policy makers continue to debate how to accurately assess social emotional learning competencies across countries (Care & Andersen 2016). Our measure of

Agency offers a versatile way to do this. By centring the voices of local organizations, we were able to develop a measure that takes into account the nuances of each community involved, and thus is applicable in many countries, as evidenced by statistical validation. This method should be instructive for many sectors beyond education seeking to ensure accuracy of evaluation and ability to capture the importance of the diverse range of approaches offered by community-driven organizations.

Bibliography:

Babyak, Michael A., Snyder, C.R., and Yoshinobu, Lauren. (1993). Psychometric Properties of the Hope Scale: A Confirmatory Factor Analysis. *Journal of Research in Personality*, 27(3), 154–69.

Bandura, Albert. (2009). Cultivate Self-Efficacy for Personal and Organizational Effectiveness--Indispensible Knowledge for Evidence-Based Management. In Edwin A. Locke (Ed.) *Handbook of Principles of Organizational Behavior*. West Sussex, UK: John Wiley & Sons Ltd.

Bandura, Albert. (1989). Human Agency in Social Cognitive Theory. *American Psychologist*, 44 (9): 1175–84.

Burton, Laura J, and Stephanie M Mazerolle. (2011). Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research. *Athletic Training Education Journal* 6, (1): 27–35.

CARE. Gender Equitable-Men's Scale. Accessed April 7, 2015: http://pqdl.care.org/gendertoolkit/Pages/Gender%20Equality%20Attitudes-Values%20Clarification.aspx

Catalano, R.F., Berglund, M.L., Ryan, J.A.M., Lonczak, H. & Hawkins, J.D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Social and Political Science, 591*: 98-124.

Cottrell, Stella. (2017). *Critical Thinking Skills: Effective Analysis, Argument and Reflection*. Macmillan International Higher Education.

Cronbach, Lee J., and Meehl, Paul E. (July 1955). Construct Validity in Psychological Tests. *Psychological Bulletin* 52(2): 281–302.

DuBois, D.L., Portillo, N., Rhodes, J.E., Silverthorn, N. & Valentine, J.C. (2011). How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological science in the public interest, 12*(2), 57-91.

Duerden, M., Witt, P., Fernandez, M., Bryant, M. & Thierault, D. (2012). Measuring life skills: standardizing the assessment of youth development indicators. *Journal of Youth Development, 7*(1). doi 10.5195/JYD.2012.155

Duckworth, Angela Lee, and Patrick D Quinn. (2009). Development and Validation of the Short Grit Scale (GRIT–S). *Journal of Personality Assessment* 91(2): 166–74.

Giddens, A. (1984). *The Constitution of Society. Outline of the Theory of Structuration.* Berkeley, CA: University of California Press.

Hillenbrand, Emily, et al. (2015). "Measuring gender-transformative change: a review of literature and promising practices," CGIAR Research Program on Aquatic Agriculture Systems, CARE USA.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change, 30*(3), 435-464.

Kautz, Tim, James J. Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. (2014). "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime Success." NBER Working Paper, National Bureau of Economic Research. http://www.nber.org/papers/w20749.

Kwauk, Christina, and Amanda Braga. (2017). "Translating Competencies to Empowered Action: A Framework for Linking Girls' Life Skills to Social Change." Washington, DC: The Brookings Institution.

Learning Metrics Task Force. (2013). "Toward Universal Learning: Recommendations from the Learning Metrics Task Force Summary Report." Washington, DC: UNESCO & The Brookings Institution.

Lerner, Richard M., Theokas, Christina and Jelicic, Helena. (2005). "Youth as Active Agents in Their Own Positive Development: A Developmental Systems Perspective." In W. Greve, K. Rothermund, & D. Wentura (Ed.s), *The adaptive self: Personal continuity and intentional self-development*, 31-47. Ashland, OH: Hogrefe & Huber.

Little, Todd D., C. R. Snyder, and Michael Wehmeyer. (2006). "The Agentic Self: On the Nature and Origins of Personal Agency Across the Lifespan." In D. K. Mroczek & T. D. Little (Eds) *The Handbook of Personal Development*, 61-79. Mahwah, NJ: Lawrence Erlbaum Associates.

Lopez, Amy, Jamie R Yoder, Daniel Brisson, Stephanie Lechuga-Pena, and Jeffrey M Jenson. (2015). Development and Validation of a Positive Youth Development Measure: The Bridge-Positive Youth Development. *Research on Social Work Practice*. 25, (6): 726–36.

Malhotra, A. & Shuler, S.R. (2005). Women's empowerment as a variable in international development. In D. Narayan (Ed.), *Measuring empowerment: cross-disciplinary perspectives*, 71-88. Washington, D.C.: The World Bank.

Maslak, Mary Ann, ed. (2008). *The Structure and Agency of Women's Education*. Albany, NY: State University of New York Press.

Mincemoyer, Claudia C, and Daniel F Perkins. (2005). "Measuring the Impact of Youth Development Programs: A National Online Youth Life Skills Evaluation System," *The Forum for Family and Consumer Issues*, 10(2):1-7.

Mulcahy-Dunn, A., King, S., Nordstrum, L., Newton, E. & Batchelder, K. (2018). The relationship between grit, self-control, and early-grade reading: a trial of measuring soft skills in rural Tanzania. *Educational Psychology, 38(*8): 997-1009.

Murphy-Graham, E. and Lloyd, C. (2016). Empowering adolescent girls in developing countries: The potential role of education. *Policy Futures in Education,* 14 (5): 556 - 577.

Nunnally JC, Bernstein IH. (1994). *Psychometric Theory,* 3rd edition. New York: McGraw-Hill.

Narayan, Deepa, Ed. (2005). (2005). "Measuring Empowerment: Cross-Disciplinary Perspectives," The International Bank for Reconstruction and Development/The World Bank. Washington, DC: The World Bank Group.

Nelson, Nici and Wright, Susan (1995). "1: Participation and power" from Nelson, Nici and Wright, Susan, *Power and participatory development: theory and practice* pp. 1-18, ITDG Publishing.

Newton, J., Eerdewijk, A. van., and Wong, F (2019). "What do participatory approaches have to offer the measurement of empowerment of women and girls." KIT Working Paper 2019:01. Royal Tropical Institute.

O'Rourke, M., Hammond, S., O'Sullivan, D., Staunton, C. & O'Brien, S. (2016). The LifeMatters programme for developing life-skills in children: an evaluation. *International Journal of Mentoring and Coaching in Education, 5*(2), 144-157.

Parrado, E., McQuiston, C., and Flippen, C. (2005). Participatory Survey Research. Integrating Community Collaboration and Quantitative Methods for the Study of Gender and HIV Risks Among Hispanic Migrants. *Sociological Methods & Research* Vol. 34(2): 204-239.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.
Scheier, Michael F, and Charles S Carver. (1985). The Self-Consciousness Scale: A Revised Version for Use with General Populations 1. *Journal of Applied Social Psychology* 15(8): 687–99.
Schermelleh-Engel, Karin, Helfried Moosbrugger, and Hans Müller. (2003) Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, Vol. 8(2): 23–74.
Schwarzer, R., and M. Jerusalem. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, and M. Johnston (Eds.) *Measures in Health Psychology: A User's Portfolio*. Windsor UK: NFERNELSON.
Schwarzer, Ralf, Judith Bäßler, Patricia Kwiatek, Kerstin Schröder, and Jian Xin Zhang. (1997). The Assessment of Optimistic Self-beliefs: Comparison of the German, Spanish, and Chinese Versions of the General Self-efficacy Scale. *Applied Psychology* 46(1): 69–88.
Sidle, Aubryn Allyn. (2019). Action on Agency: A Theoretical Framework for Designing and Operationalizing Agency in Girls' Life Skills Programs. *Gendered Perspectives in Development Working Paper Series*, #313 Michigan State University.
Sperling, Gene B., and Rebecca Winthrop. (2016). *What Works in Girls' Education: Evidence for the World's Best Investment*. Washington, D.C: Brookings Institution Press.
Torre, Dorothy Ann. (1986). *Empowerment : Structured Conceptualization and Instrument Development*. Cornell University Press.
Wheeler, M.E., Keller, T.E. & DuBois, D.L. (2010). Review of three recent randomized trials of school-based mentoring: making sense of mixed findings. *Society for Research in Child Development, 24*(3), 1-21.
Winthrop, Rebecca, and Eileen McGivney. (2016). "Skills for a Changing World: Advancing Quality Learning for Vibrant Societies." Washington, DC: The Brookings Institution, Center for Universal Education.

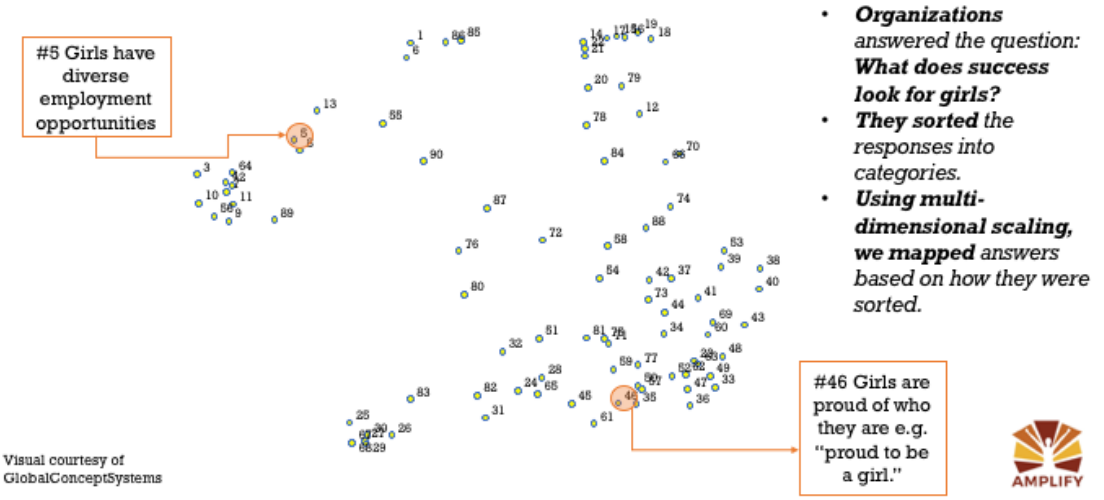
Appendix 1: AMPLIFY Members

Organization	Country	Location	Program
Action Foundation	Kenya	Urban	○ Provides essential services to children with disabilities, their families and care providers
Girls To Lead Africa	Uganda	Rural	○ Ensures that girls have equal opportunity in politics and leadership positions in their country. Program teaches girls policy making, leadership, and encourages them to participate in student politics in their schools.
AfricAid Dandelion Africa Girl Up Initiative Uganda	Tanzania Kenya Uganda	Rural/Urban Rural Urban	○ Utilizes mentorship and life skills education programs as a pathway to develop girls as leaders to transform their communities.
WISER Girls Gashora Girls Maranyundo Girls School Sega Girls The Girls' Foundation of Tanzania	Kenya Rwanda Rwanda Tanzania Tanzania	Rural Rural Rural Rural Urban	○ Provide holistic education to girls by running in-residence Boarding Schools or other in-residence programs with a focus on science and technology, and training girls as the next generation of leaders.
Chalbi Scholars Riley Orton Foundation Malkia Foundation Komera	Kenya Kenya Kenya Rwanda	Rural Rural Rural Rural	○ Holistic programs that provide access to education through scholarships, life skills education and other supports. Organizations have a wide array of other services for girl & their families.
Jifundishe	Tanzania	Rural	○ Focused on providing access to secondary education to those who have dropped out of the formal system. Supplemental programs include SRH, life skills and community engagement.
AkiraChix Elohim Development Association Streets Ahead Children's Center Association	Kenya Uganda Rwanda	Urban Urban Rural	○ Provides practical, hands on skills training that prepares vulnerable young women for existing opportunities in the labor market.

Appendix 2: Establishing Shared Metrics—Group Concept Mapping Outcomes

Group Concept Mapping: Creating AMPLIFY's Shared Metrics

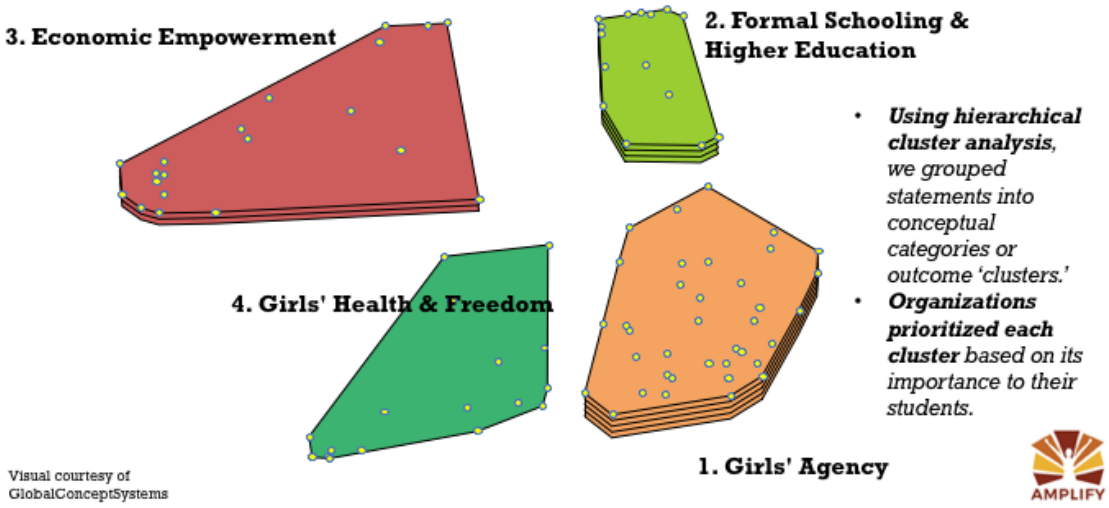
Step 1: Mapping what success looks like for Girls



- ***Organizations*** *answered the question:* ***What does success look for girls?***
- ***They sorted*** *the responses into categories.*
- ***Using multi-dimensional scaling, we mapped*** *answers based on how they were sorted.*

CREATING AMPLIFY SHARED METRICS

STEP 2: CREATING A GROUP CONCEPT MAP OF 'OUTCOME AREAS'



- ***Using hierarchical cluster analysis****, we grouped statements into conceptual categories or outcome 'clusters.'*
- ***Organizations prioritized each cluster*** *based on its importance to their students.*

Appendix 3: Full Structural Equation Model for validation of the Sidle-AMPLIFY Agency Measure