Ocean & Coastal Management
May 2016, Volume 124 Pages 53-65
https://doi.org/10.1016/j.ocecoaman.2016.02.004
https://archimer.ifremer.fr/doc/00636/74830/

Archimer
https://archimer.ifremer.fr

Understanding vulnerability in Alaska fishing communities: A validation methodology for rapid assessment of indices related to well-being

Himes-Cornell Amber [1, 2, *], Maguire Conor [3], Kasperski Stephen [1], Hoelting Kristin [4], Pollnac Richard [5]

[1] NOAA, Natl Marine Fisheries Serv, Alaska Fisheries Sci Ctr, 7600 Sand Point Way NE, Seattle, WA 98115 USA.
[2] Univ Bretagne Occidentale, AMURE LABEX IUEM, 12 Rue Kergoat,CS 93837, F-29238 Brest 3, France.
[3] Pacific States Marine Fisheries Commiss, 205 SE Spokane St,Suite 100, Portland, OR 97202 USA.
[4] Colorado State Univ, Human Dimens Nat Resources, 1480 Campus Delivery, Ft Collins, CO 80523 USA.
[5] Univ Rhode Isl, Dept Marine Affairs, Kingston, RI 02881 USA.

* Corresponding author : Amber Himes-Cornell, email address : amber.himes-cornell@noaa.gov

conormag60@gmail.com ; stephen.kasperski@noaa.gov ; kristin.hoelting@colostate.edu ; pollnacrb@gmail.com

Highlights

► Vulnerability indices for Alaskan communities are created and tested for construct and external validity. ► Our rapid assessment method develops a qualitative comparison measure to test construct validity of quantitative indices. ► Groundtruthing allows researchers to calibrate indices based on secondary data.

Keywords : Social indices, Validity, Groundtruthing, Well-being, Alaska fishing communities

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher-authenticated version is available on the publisher Web site.

1. Introduction

Historically, fishery managers placed little emphasis on studying social phenomena, opting for greater focus on biophysical and ecological disciplines. This has changed with improved understanding of the pivotal role of humans in fisheries and development of the concept of fisheries social-ecological systems (Ban et al., 2013, Clay and McGoodwin, 1995; Colburn et al., 2006; Himes-Cornell and Hoelting, 2015; Jentoft, 2006). This recognition of fisheries as complex social-ecological systems has led to efforts to understand social vulnerability of place-based fishing communities. Through enhanced understanding of conditions contributing to vulnerability, fisheries managers can better project how communities may react to perturbations resulting from policy decisions.

However, studying and reporting on fishing community vulnerability has proven challenging for social scientists (e.g., Allison et al. 2009; Boyd and Charles, 2006; Charles et al., 2009; Reed et al. 2006). Fully understanding processes affecting community resilience has traditionally required ethnographic methods. However, qualitative findings are often not well-suited to integration with standard quantitative metrics utilized in fisheries management (Sepez et al., 2006). In addition to issues of data integration, there are challenges of scale and feasibility as conducting lengthy and rigorous ethnographic fieldwork becomes increasingly resource intensive and is often precluded by demand for expedience (Jacob et al., 2010; Sepez et al., 2006).

In response to these challenges, there has been a recent effort within the U.S. National Marine Fisheries Service (NMFS) to develop quantitative indices related to community well-being derived from secondary data. Specifically, the indices measure components of community vulnerability that are theoretically linked to the larger construct of objective well-being. This effort is driven by the need to satisfy management directives outlined under the MSA while also addressing issues of data standards and timeliness. The primary goal of this national project is to create a reliable and consistent method of quantifying these constructs that remains grounded and relevant at a community level to enhance internal validity. This paper begins the process of assessing the construct and external validity of those resulting measures as well as their construct reliability (Himes-Cornell and Kasperski, 2015, 2016; Jacob et al., 2010; Jepson and Colburn, 2013).

Index validation is a critical step that must take place prior to adoption of an index into decision-making or trend analysis. Ultimately, quantitative indices are only as good as the data used to create them, and whether those data provide a valid representation of the theoretical construct the index is intended to measure. Further, it is critical to assess the degree to which construct validity is generalizable. However, although the use of quantitative indices has been growing in popularity over the last decade, relatively few studies have gone the next step to validate the results. Those that have, for the most part, still remain focused on using secondary data and analysis to undertake any evaluation (e.g., Cloquell-Ballestar et al., 2006; Fekete, 2009; Lyubomirsky and Lepper, 1997; Sherrieb et al., 2010; Tate, 2012; Tate, 2013).

We argue that a more effective method for testing index validity is to gather ethnographic data that can be used to “groundtruth” (Smith et al., 2011) the quantitative indices against the real world. Methodologically, we argue that comparing qualitative, ethnographic data for a representative sub-set of communities to their respective quantitative index rankings allows the researcher to test for convergence. If the two measures are highly correlated, it provides evidence that the quantitative well-being indices possess a sufficient level of construct validity to justify their use in policy and planning processes. In other words, it presents evidence that the quantitative indices are oriented in reality, rather than being a product of the methodology itself (Johnson et al., 2007); convergence of the two measures demonstrates that the quantitative indices, and the secondary data on which they rely, accurately reflect real-world conditions found in sampled communities.

This paper reviews development of these quantitative well-being indices (Himes-Cornell and Kasperski, 2016), the creation of a community typology using cluster analysis, and ethnographic fieldwork that was conducted to develop an independent "qualitative comparison measure" of well-being for a sub-set of communities identified as possessing distinct characteristics within the typology (e.g., Jepson and Jacob, 2007; Smith et al. 2011). We present a multi-step methodology for and the results of a rapid, qualitative assessment of the construct and external validity of the quantitative indices. This methodology is inspired by similar work carried out in fishing communities in the Gulf of California (Morzaria-Luna et al., 2013), Gulf of Mexico (Jacob et al., 2010 and 2013), New England (Colburn and Jepson, 2012), northern Australia (Marshall and Marshall, 2007), and Puget Sound (Biedenweg et al., 2014).

Although this paper focuses on a methodology used to test the validity of quantitative indices specific to fishing communities, the primary goal is to develop a rapid assessment methodology that can be used to test the validity of other types of indices or indicators based on secondary data that are used in other fields of research, not just in the case of fisheries as described here. Our results highlight numerous obstacles to development of valid quantitative well-being indices from secondary data. These challenges are associated with data reliability questions arising from field logistics as well as quantitative data quality issues. However, despite these obstacles we believe quantitative well-being indices remain a promising and useful method that can be used to fulfill an important management need. Moreover, we believe that an index validation methodology such as the one presented in this paper can be viewed as a first step in the validation process, where we identify which indices and constructs need refinement. This step can assist in identifying and mitigating problems related to data quality and field logistics and can be followed up with additional groundtruthing steps to create an iterative validation approach.

2. Material and methods

2.1. Definitions

Understanding vulnerability to economic, social, and environmental instability accomplishes an important step in assessing how fishing communities may respond to disturbances, and may contribute to better tools for making institutions more adaptive and robust. Here we follow the MSA definition of “fishing communities” as a status of communities which depend significantly on fish harvesting or processing to meet social and economic needs (MSA, 2007). We define fisheries engagement as the extent to which a community is represented within aggregate fishing activity across all fisheries in which its residents participate. Fishing dependence is a more local concept, reflecting per capita involvement of local residents in fishing activities, and is a measure of how important fishing is to the health of the local economy (Himes-Cornell et al., 2013). While this serves a purpose in terms of creating an operational definition of “fishery dependence,” it does not address the cultural and social values inherent in that term (Brookfield et al., 2005). To those living in a community, fishery dependence may not be limited to reported landings and other associated fishing activity (e.g., vessels owned or fishing permits held by local residents), but may be inextricably linked to their cultural connection to the act and ritual of fishing. It is important to explore these concepts if managers are to better understand the structure and needs of fishery-dependent communities, as well as how they react to changes in their social-ecological environment.

Many researchers can attest to the difficulty of quantifying concepts such as vulnerability, resilience, and well-being (Allison et al., 2009; Boyd and Charles, 2006; Reed et al., 2006). For the purposes of this study, we focus on the general definition of well-being provided by Pollnac et al. (2006). It incorporates both subjective and objective well-being, and is situated within the context of fisheries social impact assessment: "Well-being refers to the degree to which an individual, family, or larger social grouping (e.g. firm, community) can be characterized as being healthy (sound and functional), happy, and prosperous" (p. 2).

2.2. Index Validity - Evidential and Consequential

Criticisms of social indices are often associated with evidential validity. Evidential validity is concerned with both measurement validity (the degree to which a measuring instrument succeeds in measuring a theoretical construct, including construct validity, criterion validity, and content validity) and the validity of causal inference (internal and external validity) (Adcock and Collier, 2001). Each of these aspects of evidential validity addresses different links in a "chain of evidence-based inferences" (Guhn et al. 2011, p. 186). As an initial step in index validation, the methodology presented in this paper is specifically geared toward assessment of construct validity (testing for convergence between two theoretically related measures) and external validity (testing for consistency in convergence across communities).

Construct validity, similar to the overall concept of evidential validity, can be described as the degree to which an observed measure (e.g., an index) accurately reflects the theoretical construct it is intended to measure (e.g., Adcock and Collier, 2001; Andrews and Withey, 1976;

Connidis, 1984). The construct validity of an index can be compromised by a variety of factors in its development, such as poor internal validity (the quality of theoretical relationships among indicator variables, and cause and effect relationships between variables and the construct), poor content validity (ensuring that all theoretical elements of the construct are adequately represented by selected indicator variables), or poor construct reliability, whether due to poor secondary data quality or low inter-observer agreement, etc.

Two common techniques for assessing the level of construct validity possessed by an index are 1) to test for convergence with another measure of the same construct that is expected to be highly correlated (a.k.a. convergent validity), and 2) to test for divergence with a measure that in theory would not be expected to be correlated (a.k.a. divergent, or discriminant, validity) (Adcock and Collier, 2001; Guhn et al., 2011).[1] The assessment of construct validity presented in this paper specifically tests for convergence between two independent measures of community well-being that are expected to be highly correlated, given their intent to measure the same theoretical construct. We refer to this as a test of convergent construct validity. Results of this construct validity assessment can help determine which index components already possess high construct validity, and which may require additional attention to underlying issues of internal and content validity, as well as construct reliability issues. Building off of Jacob et al.'s (2013) attempt at validating quantitative indices, the methods presented in this paper also provide an initial assessment of construct reliability through a test of inter-observer agreement in the development of the qualitative comparison measure.

---

[1] A third method for assessing construct validity is to test for criterion validity, which assesses how well a measure's scores correlate with the scores of an accepted indicator, or "criterion" variable (Adcock and Collier, 2001, p. 537). The validation method presented in this paper uses a test of convergence between two theoretically related constructs, as opposed to a comparison of index scores against the scores of an established criterion variable (Guhn et al., 2011, p. 185).

External validity has to do with whether the index is generalizable, in this case across distinct types of communities. It is possible that the variables used to develop an index may adequately represent the characteristics of one community, but fail to represent other characteristics present in another community. These discrepancies can only be identified by testing for construct validity across multiple community types. The methodology presented in this paper begins to address the question of external validity by examining how convergent construct validity and construct reliability (inter-observer agreement) vary across communities found to possess distinct characteristics. It is important to note that the methods and results presented here are a preliminary assessment of external validity and will be expanded on in future work on this project.

Beyond assessing the evidential validity of an index, it is also critical to consider the validity of index application. Validity has to do not only with measurement and theoretical grounding, but also "the interpretations, uses, and consequences that are based on measurement scores and that ensue from the measurement process" (Guhn et al. 2011, p. 184). This issue, increasingly referred to as consequential validity, has to do with the potential distributional ramifications that may result from use of an index in decision-making. It highlights the fact that application of an index may not be appropriate in all contexts and for all purposes (Guhn et al., 2011; Messick, 1998). The methods presented in this paper do not directly address consequential validity, but the authors acknowledge the need to carefully assess when and how application of the objective well-being indices would be useful, appropriate, and considered valid by affected populations.

2.3 Methods

This paper presents a multi-stage methodology used to first develop a set of qualitative indices of community well-being, and subsequently to test for construct and external validity of the resulting indices. We use a mixed-methods approach (Creswell, 2003; Creswell et al., 2011) applying quantitative methods (i.e., construction of quantitative indices) and qualitative methods (i.e., ethnographic data collection; grounded theory). The approach involved seven steps, which are outlined here in chronological order. The methods involved in each step are presented in more detail below.

1) Construction of quantitative indices of well-being using principle component factor analysis (PCFA);
2) Development of a community typology using cluster analysis, which uses the index values developed in Step 1 to group communities with others that possess similar characteristics;
3) Completed ethnographic groundtruthing fieldwork in communities selected to represent distinct groupings from the community typology, with the aim of developing a test of the external validity of the quantitative indices;
4) Development of a qualitative comparison measure using interview and observational field data to assign subjective ranks to categories that matched index components (factors) identified in the PCFA in Step 1;
5) Statistical assessment of construct reliability using researchers' subjective rankings to test for consistency across communities to ensure reliability of the qualitative comparison measures;
6) Development of a comparable ranking system for quantitative index components to match the qualitative ranks described in Step 4; and

7) Statistical assessment of convergence between qualitative rankings and quantitative indices to test for construct validity of the quantitative indices.

Confidence in the results of the convergent construct validity tests relies on two assumptions: 1) the ontological assumption that there is a measurable objective reality that is dictated by interactions of actors within their SES (Charmaz, 2008); and 2) that our observations of that reality are more accurate than index conclusions. While quantitative data is objective in that is has been standardized and strictly defined, our observations, and those of interviewees, are grounded in subjective experience (Mills et al., 2006). This can lead to struggles when reconciling qualitative and quantitative data. However, verification of the reliability of qualitative observations via inter-rater agreement tests, such as the one used in Step 5, help increase confidence that those observations are grounded in reality as long as we accept that multiple descriptions of phenomena can exist without being in contradiction (Heath and Cowley, 2004).

2.1. Step 1: Quantitative indicator development

The first step in our methodology involved development of a set of quantitative indices using variables that represent distinct components of the overarching well-being construct. Variable selection was guided by work on social vulnerability to environmental hazards from Cutter et al. (2003) as well as similar work on fishing community vulnerability by Colburn and Jepson (2012) and Jepson and Colburn (2013). In addition, to modify Colburn and Jepson's methodology, we added variables to capture unique characteristics of vulnerability and well-being in Alaskan communities. In many cases data were highly skewed, in which case we employed a $\log_{10}$ transformation to make patterns more apparent. Summarized in the Appendix (Tables A1 and A2), the full data set includes 78 social and 73 fisheries variables collected for

346 Alaska communities (determined as Census Designated Places). Due to missing data for a number of communities, our resulting indices only included 284 communities throughout the state of Alaska.

The variables were drawn from a variety of state and federal sources, using average values over the period of 2005-2009. Social and economic data were compiled from sources including U.S. Census Bureau 2000 Census (U.S. Census Bureau 2000) and 2005-2009 5-year estimates (U.S. Census Bureau, 2010a, b), the Alaska Local and Regional Information Network (ADLWD, 2001 and n.d.), education statistics and reports (Alaska Department of Education and Early Development, n.d.), Community Database Online (Alaska Department of Commerce, Community, and Economic Development, n.d.), and various other sources (Himes-Cornell et al., 2013; Himes-Cornell and Kent 2013a, b). Fishery data were compiled by the Alaska Fisheries Information Network (AKFIN, n.d.) drawing from sources including the National Marine Fisheries Service (2011a-d), Alaska Department of Fish and Game (ADF&G, 2011a-c: Fall and Koster, 2011; Fall et al. 2011), Alaska Commercial Fisheries Entry Commission (CFEC, 2011, 2015), the U.S. Fish and Wildlife Service (USFWS, 2011) and the Alaska Beluga Whale Commission (Frost and Suydam, 2010).

Given the large number of variables identified, we elected to use principal components factor analysis (PCFA), a data reduction technique, to reduce them to a manageable level and to identify latent index components that serve as measures of distinct elements of well-being. We conducted separate PCFAs first using social data (e.g., poverty, employment), and then fishery data (e.g., landings, permits). We used a scree test to determine the number of components that could be considered in the PCFA, where the number of components appropriate to consider corresponded to the inflection point of the scree plot. During this step, we used a varimax rotation of the factor loadings with Kaiser normalization in order to isolate variables that have the highest factor loading for each component. This was meant to ease interpretation of factor loadings by altering them so that they were more discretely attributed to each factor. Quantitative well-being index scores for each of the components of well-being were constructed using the regression method and are normalized to have a mean of zero and standard deviation of one. An Armor's theta reliability test was used in order to test the internal consistency of the variables in each component, where a value of theta greater than 0.5 is considered acceptable (Jepson and Colburn, 2013; Himes-Cornell and Kasperski, 2015, 2016; Smith et al., 2011). Ultimately, the final analysis was able to maintain theta reliability scores above 0.8; confirming the reliability of the PCFA instrument.

Quantitative indices such as those presented here in Step 1 are only useful as long as they exhibit an acceptable amount of construct validity, meaning how well the indices represent the communities they measure (Jacob et al., 2013). While individual variables affecting vulnerability and well-being can often be quantified, producing a reliable composite index presents more of a challenge. Interaction between variables and how they collectively contribute to overall well-being is poorly understood, making it difficult to understand their influence on overall community well-being and vulnerability (Kelly and Adger, 2000). Moreover, it is difficult to determine the generalizations can be made from context-driven variables or how the insights gained can help explain how perturbations affect individual communities (Boyd and Charles, 2006). Because of this, groundtruthing is an important next step in validating the representativeness of indices as well as formulate a context in which to apply them.

2.2. Step 2: Cluster Analysis to Generate a Community Typology

To begin the groundtruthing process, we used a cluster analysis to group communities into a typology based on the results of the two PCFAs conducted in Step 1. Identification of community types was important because it allowed us to design our ethnographic data collection in such a way that we could begin to assess external validity of the quantitative indices. It also served the practical purpose of reducing the number of communities we would need to visit in order to capture differences among communities (Smith et al. 2011).

To develop the typology, we used a non-hierarchical K-means cluster analysis technique to group multivariate data through a process of maximizing between-group variability, while minimizing within-group variability (Smith et al. 2011). The clustering process used component scores derived from the transformed variables used in both the fishery and social PCFAs. Communities were then grouped into a fixed number of predetermined clusters. This was accomplished by analyzing overall Euclidian distance from an empirical mean of all cases (communities) and creating "seeds" based on the number of clusters desired. Seeds selected are as far as possible from the center of all the cases. Communities were then assigned to their nearest seed and then reassigned if necessary to reduce within group sum of squares, minimizing within-group variability (Jain, 2010; Smith et al., 2011).

Several exploratory cluster analyses were conducted using 7, 15, 20, 25, 30, and 35 clusters. The goal was to determine an appropriate number of clusters that accurately grouped communities based on our knowledge of Alaska's communities. We examined the PCFAs component scores in conjunction with the cluster analyses to gather a better picture of what characterized each cluster. In this case a higher index score equated to a higher influence of a particular component, and vice versa. Finding a balanced number of clusters proved challenging, as a smaller number of large clusters risked grouping communities that should not be together, while a large number of smaller clusters could overly disperse communities, impacting their usefulness. The decision of the number of clusters to create in the analysis was reached by comparing each iteration of the cluster analysis (i.e., 7, 15, 20, 25, 30 and 35 clusters), and determining whether communities fit in their respective clusters based on a review of available literature on community characteristics, community profiles (Himes-Cornell et al., 2013), and original (untransformed) social and fishery variables (e.g., grouping known large multi-species commercial fishing communities together). A degree of researcher interpretation was necessary to determine if there were any glaring errors in delineations, which might reveal data errors. Ultimately, we decided that an analysis based on the creation of 25 clusters was most appropriate and useful (see Table A3).

We identified at least one community from each cluster that was influenced by fishing activity for the qualitative fieldwork phase of this research. Sample site selection was determined according to cluster representation, as well as time and budget constraints. An attempt was made to conduct fieldwork in as many communities as possible by focusing on communities that spanned all of the clusters but were located within a feasible geographic range. Each cluster was analyzed to determine which communities were both geographically close to each other, and the most central in (or representative of) the cluster (as determined by Euclidean distance from its center). Ultimately, we selected a total of 13 communities for the fieldwork component, representing 11 of the 25 clusters.

2.3. Step 3: Field-based groundtruthing

We developed an ethnographic fieldwork protocol using a multifaceted grounded theory approach. First, a stakeholder analysis was required to identify key informant categories to target for interviews (Prell et al., 2009; Reed et al., 2006). For each community selected for fieldwork, we gathered historic and contextual information as a starting point (Himes-Cornell et al., 2013). This information was independent of the secondary data used in the creation of the quantitative indices, and was based on a comprehensive search of available literature. Through this, we identified expected informant types for each community, including community leaders; commercial, recreational, and subsistence fishermen; fishery support businesses; and other local businesses and services. Selected informant types were then compared with relevant aspects presented in the component scores of the PCFAs in order to confirm that their expertise was relevant to variables that were thought to be heavily influential to the community.

Once informant types were identified, interview topics were chosen so that we could undertake fieldwork while possessing an understanding of salient themes with which to best engage respondents. Available literature was referenced against the PCFA components to identify themes that could be used as interview prompts. Recognizing the potential for bias in the initial selection of interview topics, we included an iterative, soft systems approach to identifying additional topics while in the field (Reed et al., 2006; Mingers, 1980). Allowing informant-identified topics to emerge during the interview process and using them to further inform the interview process going forward helped correct misinterpretations of community character and well-being that may have biased initial selection of the interview topics.

The initial interview topics were adapted into a field protocol that guided open-ended interviews. Topics were categorized into specific key-informant protocols based on unique characteristics of groupings of informants, including commercial fishermen, recreational fishermen, subsistence fishermen, local business owners, and community leaders. In addition, we developed a general protocol that included topics to discuss in all interviews. Interview topics are summarized in Table 1.

Interviewers were allowed a large degree of latitude when determining the flow and content of the interview. In many cases, informants were allowed to determine the direction of the interview while the interviewer posed topics ensuring that discussions addressed themes pertaining to targeted constructs and the informant's relationship with them. As the fieldwork team became more familiar with locally salient themes, they became more adept at gathering thematically targeted perspectives while continuing to build from them. This allowed interviewers to target core themes, while continuing to use broad themes so that each informant had an opportunity to identify new ones.

2.3.1. Conducting ethnographic fieldwork

Fieldwork was divided into three segments that took place between May and September 2013, with each trip lasting between 10 and 16 days. Time spent in each community was determined according to population, with larger communities receiving longer visits. Effort was made to contact key informants prior to arrival so that we would be able to become quickly oriented with fieldwork sites upon arrival. We used random sampling, purposive quota sampling and snowball sampling methods to ensure a broad spectrum of informant types were interviewed. We asked each informant interviewed through the random and purposive quota sampling techniques to recommend additional community members who would be able to provide a useful perspective.

A total of 286 (n = 286) informants were interviewed across communities; a summary of interviews can be found in Table 2. Several protocols were administered in situations where a single informant satisfied multiple roles, resulting in an interview protocol tally exceeding the total number of informants (Table 2). Determining adequate sample size was dependent on the community being studied. For larger communities (N > 200), we attempted to interview 20-30 informants, while 10-20 interviews were attempted in communities with populations less than 200 (N < 200). These targets also allowed us to achieve content saturation, as well as take a pragmatic view of what could be accomplished under time and resource constraints. In a review of available literature, Mason (2010) highlights the diverse opinions regarding adequate sample size, ranging from a minimum of 15 respondents, to a maximum of 30-50 for grounded theory applications. However, a range of influences affected how many interviews were attained in addition to population size. These included the availability of venues, weather, timing, community layout, and the willingness of residents to participate. Thus, in the tradition of mixed-methods pragmatism, a flexible sampling method was adopted that responded to conditions present in sample sites (Giddings and Grant, 2007).

During fieldwork, an effort was also made to assess physical assets and characteristics of a community. This included an inventory of available services and infrastructure as well as a photo survey. Some elements of community infrastructure were included in the original dataset; however, the ground assessment aided in validating data and improving quality. Photo surveys targeted elements of the community that we thought to be unique or important to its character. These included culturally defining elements (e.g. locally produced artwork, landmarks), community style or aesthetics (e.g. community centers, unique or defining architecture), fisheries-related infrastructure (e.g. harbors, docks, seafood processors), physical landscape (e.g. natural spaces, topography), and other elements that helped characterize the community (e.g. community message boards). In addition to informing and supplementing data, photo accounts aided us in assessing the overall physical condition of the community. Finally, workshops were held in communities where interest was expressed. In addition to familiarizing community members with the research, these workshops provided an opportunity to collectively discuss and refine the interview topics.

2.4. Step 4: Development of a qualitative comparison measure

Field visits to the 13 communities involved either two or three independent researchers. Following completion of the field visit, each researcher was asked to assign subjective rankings for each community based on interviews and personal observations in the community. Subjective rankings were based on categories matching the individual index components (factors) identified through the two PCFAs (see the first column of Tables 4 and 5 for these factors). The magnitude of these ranks was categorized and coded numerically as follows: “high”=3, “medium”=2 and “low”=1. For example, if a team member perceived that a community had high levels of poverty (e.g., high unemployment, poor living conditions), than he or she would assign a rank of 3 to the corresponding “poverty” construct, and so on.

2.5. Step 5: Statistical assessment of construct reliability

The qualitative ranking method described in Step 4 generated two or three independent ranks per qualitative category per community, depending on the size of the research team during the field visit to each community. We tested for the consistency of these rankings using an inter-rater agreement test in order to determine the level of construct reliability of the qualitative comparison measure.

Inter-rater agreement is commonly assessed using one of the following statistical tests: percentage agreement, correlation statistics (e.g., Pearson’s *r*, Spearman’s rho), or Cohen’s kappa. Following Jacob et al. (2010, 2013), we selected a weighted Cohen's kappa statistic $(\hat{\kappa})$ to measure the degree of consistency between the qualitative ranks of multiple team members (Cohen, 1960, 1968). This was chosen over a simple percent agreement because it produces a more conservative measurement by adjusting for agreement due to random chance. Weights were assigned depending on how far apart team members' ranks were, with less weight given to pairings that were farther apart. Rather than simply testing for perfect agreement, this allowed us to incorporate a degree of agreement which is useful when considering the subjective nature of qualitative ranking (Viera and Garret, 2005).

The weighted Cohen's kappa statistic comparing two individual raters (referred to above as team members) is calculated by taking percentage of observed agreement (P_a) and subtracting expected random chance agreement (P_e), divided by 1 minus expected random chance agreement, such that:

$$\hat{\kappa} = \frac{P_a - P_e}{1 - P_e}. \quad (1)$$

As there are three categories (k=3) that a rater can choose (high, medium, low), agreement is weighted among raters based on their strength of agreement using:

$$\omega_{ij} = 1 - \left(\frac{|i-j|}{k-1} \right) \quad (2)$$

where i and j index the scores (high=3, medium=2, low=1) for any pair of raters. Perfect agreement (e.g. high/high) was assigned a weight of 1, partial agreement (e.g. high/medium) was assigned a weight of 0.50, and poor agreement (low/high) was assigned a weight of 0. This allowed for the inclusion of partial agreements when they otherwise would have been excluded. The percentage of observed agreement is:

$$P_a = \sum_{i=1}^{k} \sum_{j=1}^{k} \omega_{ij} p_{ij} , \qquad (3)$$

where p_{ij} is the percentage of ratings i by rater 1 and j by rater 2 (Fleiss, Levin, and Paik 2003). The expected random chance agreement is:

$$P_e = \sum_{i=1}^{k} \sum_{j=1}^{k} \omega_{ij} p_{i.} p_{.j} , \qquad (4)$$

where $p_{i.} = \sum_j p_{ij}$ and $p_{.j} = \sum_i p_{ij}$.

For each community, each team member's qualitative ranks were compared against each other using this weighted kappa to produce a measure referred to as "inter-observer reliability." Since Cohen's kappa is a two-rater test, it was performed two to three times for each community depending on how many team members were at a given site. If observers were not in adequate agreement (low construct reliability), then results from the construct validity test (Step 7) for that community were determined as inconclusive due to poor reliability of qualitative observations. To be considered adequate, an average kappa statistic of at least 0.20 was required across pairs of observers. In addition, the p-value associated with the kappa statistic must be below 0.05 in order to be considered statistically significant (Viera and Garrett, 2005). Landis and Koch (1977) provide a useful scale for kappa interpretation in which a kappa statistic of 0.20 or greater signifies an acceptable amount of agreement (Table 3). With relatively few sets of observations to compare, at least one test of team member agreement had to produce statistically significant results for an average kappa statistic to be accepted and used in Step 7 (i.e., the final assessment of the quantitative indices' construct validity). Justification for this is based on the fact that with fewer observations, each observer carries more weight. For example, in cases where there were three sets of observations, one statistically significant result accounted for 66% of observations (or 2 out of 3 observers).

Finally, we tested how consistently the team members were cognitively framing each of the individual constructs across communities. If interviewers had not been cognitively framing constructs in ways that were compatible with each other or in relation to the quantitative indices, their qualitative ranks would not be commensurable. In theory, if team members were conceptualizing constructs in ways consistent with each other, then very little variation would be seen when comparing team member agreement on that construct across each community. For example, if team members A and B both agreed that poverty was low in community X, then they should be able to apply the same assessment criteria when observing conditions of poverty in community Y. However, if while in community Y, team member A assigns a rank of low, while team member B assigns a rank of high, then there is a breakdown of conceptual consistency and we must re-examine how we are framing poverty.

We conducted a construct reliability test across team members for each individual construct, as opposed to each sample community (as described above). Again, we calculated a weighted Cohen's kappa statistic based on paired ranks provided by each researcher. In this analysis, constructs were the unit of analysis instead of communities, and the same acceptance parameters were used for the kappa statistic as for the previous tests. This test allowed us to determine whether it was appropriate to perform the construct validity test in Step 7 (below). If team members were conceptualizing constructs (e.g., poverty) in ways that were incommensurable, then it may not be appropriate to use these qualitative rankings in the analysis.

2.6. Step 6: Development of a comparable ranking system for quantitative index components

Following the magnitude scale used for the qualitative constructs, the well-being component scores for each community were again ranked "high"=3, "medium"=2, or "low"=1.

While the quantitative indices are all mean zero and standard deviation one, many index scores were positively skewed; therefore, we used a Jenks natural breaks classification method to prevent a misleading number of communities assigned with "low" ranks across indices (ESRI, 2011). This method is similar to a single dimension K-means cluster analysis, assigning index scores to the three possible ranking groups based both on their magnitude and their relationship to each other.

2.7. Step 7: Statistical assessment of convergent construct validity

To assess the construct validity of the quantitative indices, we examined convergence between quantitative and qualitative rankings by measuring inter-rater agreement with a weighted Cohen's kappa test (Jacob et al., 2010; Jacob et al., 2013). Multiple two-rater weighted kappa statistics were calculated for all 13 sample communities (StataCorp, 2011). Like the inter-observer agreement (construct reliability) examined in Step 5, this test assessed the degree to which two observations converged on a single conclusion (McHugh, 2012). However, instead of measuring agreement between team member's rankings, this time the two-rater weighted kappa statistic was used to compare each team member's qualitative ranks with the communities' corresponding quantitative ranks in order to determine the degree of agreement, and thus how well the quantitative index scores reflect reality. Again, acceptable inter-observer agreement had to have been reached in Step 5 in order for this test to proceed.

As with the previous test, if at least one test result was statistically significant then the kappa statistics from each test for that community were averaged to create a single composite kappa (Conger, 1980). This averaged kappa was then compared against the Landis and Koch scale (Table 3) in order to determine the construct validity of the well-being index associated with it. This scale allowed us to determine the degree of representativeness a particular index possessed. Communities with an average kappa statistic below 0.20 or a kappa statistic that was not statistically significant (p-value > 0.05) were determined to have index scores with poor or questionable construct validity (Viera and Garrett, 2005). This method adopts a slightly different approach than the inter-observer reliability test described in Step 5, in that in statistically results do not automatically discount the construct validity test for that community. This is due to the assertion that if team members were in acceptable agreement, then their observations of reality are accurate, thus negating the difference between poor agreement and agreement due to random chance.

3. Results

3.1 Quantitative indices

Ultimately, the PCFAs conducted in Step 1 created seven components of social vulnerability explaining 62% of variance; and eight components of fishery dependence explaining 72% of variance (Tables 4 and 5). The social components were labeled as the following: community size, infrastructure, rural/village character, poverty, transient population, foreign-born/Asian population, and retirees/low female labor force participation. Fishery involvement components were then labeled as the following: fishery participation, fishery participation per capita, crab/ American Fisheries Act (AFA)/Federal Processing Permits (FPP), sportfishing participation, FPP per capita/sea otter subsistence, local landings/vessels/processors, marine mammal and salmon subsistence, and federal crab permits/beluga harvests. The social components were intended to capture a snapshot each community’s overall (objective) social well-being, while fishery involvement variables were intended to measure dependence on, and engagement in, commercial, recreation, and subsistence fishing activities. Component categories were selected based on groups that were loaded heavily toward a single factor. Construct names were then chosen to best describe the variables included in each factor.

3.2 Construct reliability test

Overall, construct reliability was fairly consistent (Table 6). Of the 19 constructs, only two were considered inconclusive ($p < 0.05$); low female workforce and salmon subsistence. Of the average kappa values that produced statistically significant results, only beluga harvesting had a kappa that fell below 0.20 and was determined to have slight agreement. By assessing these results, we can determine constructs that may warrant further investigation in terms of how we are defining them. Ultimately, constructs with slight or inconclusive agreement may impact results of the inter-observer reliability tests by confusing real world conditions with team members' personal interpretation of those conditions. Therefore, this test can act as an initial diagnostic of the overall method by highlighting differences in the cognitive processes that provide the foundation for qualitative ranking.

3.3 Comparing qualitative and quantitative rankings – convergent construct validity test

The results of the inter-observer reliability and final construct validity tests are found in Table 7. Indices for two communities, Seldovia and Dillingham, failed to produce statistically significant results in either or both of the inter-observer reliability and construct validity tests, and were given inconclusive designations. Indices for five communities, Kodiak, Naknek, Ouzinkie, Port Lions, and South Naknek, exhibited poor construct validity either due to low average kappa statistic or high probability of agreement being attributed to random chance (i.e., p-value > 0.05). Indices for six communities, Aleknagik, Kenai, King Salmon, Port Graham, Sand Point, and Soldotna, exhibited fair or higher construct validity, resulting from a statistically significant kappa statistic of 0.20 or greater.

3.4 Preliminary Assessment of External Validity

Results of the construct validity assessment were inconsistent across communities that possess distinct characteristics. This is true both of index components that exhibited high or moderate construct validity, as well as those possessing low construct validity. This suggests that the objective quantitative indices possess only a moderate degree of external validity related to how well they represent real-world conditions in distinct community types. However, this external validity assessment is not conclusive given that only 7 of the 25 distinct community types were included in groundtruthing data collection. However, the method presented here could be expanded to test across a greater spectrum of community types.

4. Discussion

The methods described here aim to establish a rapid ethnographic assessment methodology to begin to test the evidential validity of quantitative indices. Specifically, we employed tests of construct and external validity. A first step in external validation was accomplished by selecting representative communities for groundtruthing fieldwork using a community typology generated using cluster analysis. Construct validity was assessed by testing for convergence with an independent qualitative comparison measure of well-being derived from groundtruthing fieldwork.

Ultimately, the results gave a mixed impression of the validity of the indices as an attempt to provide insight into community well-being. Objective well-being is very place-specific, and it will always be a challenge to design a generalized measure. It is also a nuanced construct, and it appears that broadly applied metrics may not adequately describe conditions that are place-specific in scale. This does not necessarily negate the usefulness of the indices developed in Step 1. Application of this validation method helps us identify components that fall short when applied broadly, as well as those which work well at the place level. Moreover, this form of rapid assessment allows researchers to not only address validity concerns, but to determine conceptual or geographic areas where additional research effort is needed. This could include additional fieldwork in a community or representative cluster of communities, or modification of a particular construct so that it may provide better insight into community well-being.

During the groundtruthing process, challenges and limitations emerged throughout each phase. These limitations and caveats must be addressed in order to better understand the methodology's strengths and weaknesses. Overall, time and resources available presented the largest challenge to conducting fieldwork in each location. Depending on respondents' willingness to participate, it was sometimes difficult to build rapport when time in a community was limited. Some respondents distrusted the team member's motivations or were hesitant or unwilling to converse with us regarding subjects that they found sensitive. Others would only allow us limited access to their perspectives, sometimes cutting interviews short. While these challenges were present in most communities, they were manageable and did not inhibit our ability to conduct research in any of the sample sites. However, inconclusive results in some communities may have been due to data limitations.

The complexity of the groundtruthing process was of concern as well, and it was often challenging for two to three researchers to conduct interviews using an iterative and adaptive process while maintaining consistent interview styles, especially given the semi-structured nature of the interviewing methodology. However, this is a trade-off we wanted to make in order to take advantage of interviewee experiences that were slightly tangential to our formal interview topics. While research conditions at times were less than ideal, pragmatism dictated that research should be adaptive and flexible, working with what is available to produce the best possible results (Giddings and Grant, 2007; Glaser, 1992; Heath and Cowley, 2004).

Interpreting results from the PCFAs also produced challenges for ranking qualitative constructs. In several instances latent components that emerged in the PCFAs were influenced by redundant or seemingly unrelated variables. Because of this, some components either seemed duplicative (e.g., “crab, American Fisheries Act, and Federal Processing Permits” and “number of crab permits” constructs; refer to Table 5), or were difficult to separate from each other for the purpose of qualitative ranking or to observe during fieldwork (e.g., “retirees/low female workforce;” refer to Table 4). Interpreting factor loadings presented a unique challenge when seemingly disparate variables combined into the same component. In addition, it was difficult to categorize components into constructs in ways that would be easily discernible in the field. We dealt with these challenges by categorically separating such components into two constructs before ranking them as qualitative measures (e.g., separating “retirees” from “low female workforce”). When the time came to compare qualitative and quantitative ranks from each individual researcher, the constructs were condensed back to their original components using a simple modal response method similar to that used by Jacob et al. (2013). This conservative approach allowed for identical ranks for each construct to be preserved, while those that differed regressed to a more neutral rank. For example, if a researcher gave a ranking of “high,” or “medium” to the “retirees” qualitative measure, and ranking of “low” for the “low female workforce" qualitative measure, then the condensed qualitative rank of "medium" would be used for comparison with the quantitative component.

In terms of the construct reliability test (inter-observer agreement), constructs that tested either as not reliable or inconclusive were also among those concepts that were the hardest to distinguish based on visual inspection of the community and/or may have only been recorded as an interview topic by a single interviewer (or none at all). Identifying potential weaknesses and strengths in qualitative observations allowed us to identify which constructs may need additional framing and refining, and guide inclusion of appropriate caveats when presenting results. The presence of three inconsistently framed constructs does not discredit results of the other tests since the majority of constructs were found to be reliable. Identification of inconsistent constructs can help us improve the quantitative indices in the future. Moreover, identifying specific problematic variables in the indices provides important context when looking at construct validity because it can challenge positivist assumptions pertaining to observations, at least in relation to those specific components. Conversely, inconsistencies may reflect insufficient qualitative data, which would support additional scrutiny when developing qualitative comparison measures, as well as warrant further study into those particular conceptual areas.

For example, team member A may have given a rank of 2 to beluga subsistence in Aleknagik based on interviews with residents who described belugas traveling up the Wood River, while team member B may not have interviewed anyone who described belugas as being an important subsistence resource, thus giving a rank of 1. This shows how agreement can hinge on the quantity of interviews and emphasizes how important reaching a saturation point is for gathering reliable qualitative data. The point at which qualitative data has reached a point of saturation is often determined during the coding process (Guest et al., 2006), although it can also be assessed ad hoc while in the field. In addition, within the context of construct ranking, it can be assumed that highly salient themes have a better chance of emerging during interviews; therefore frequency and detail of those themes can be used as a barometer for relative importance in the community. Returning to the beluga subsistence example, if beluga subsistence is truly important to Aleknagik as a whole, then the chance of beluga harvesting being mentioned during interviews is increased by virtue of it being a salient theme. As long as there is an adequate sample size, then it can be determined that relative importance is tied to how often the topic is introduced. Taking into account the inter-observer reliability test described in the methods Step 5, this means that team member A's rank of 2 and team member B's rank of 1 are in fact both an accurate reflection of reality as they experienced it through their interviews (again assuming that multiple descriptions of phenomena can exist without being in contradiction).

While the inter-observer reliability test offered reassurance that constructs were mostly being framed in similar ways, it did not account for the larger issue of whether or not team members were framing constructs in ways compatible with the quantitative indices overall. This issue arises from the fact that while component scores were ranked in relation to all 284 communities used in the PCFA (Methods Step 1), the reference scale available to team members was limited only to the communities they visited. Control for this is then dependent on how representative community clusters are (Methods Step 2), as well as the number of clusters visited during fieldwork (Methods Step 3). Since only 11 out of 25 clusters were visited, these potential impacts on testing construct validity (Methods Steps 4 through 7) must be recognized.

5. Conclusion

The primary goals of this paper are 1) to develop a rapid assessment methodology for validating a specific set of quantitative indices of fishing community well-being using ethnographic data collection and 2) to present a methodology that can be used more broadly to validate quantitative indices. The groundtruthing method presented here is a rapid qualitative assessment methodology which allows for development of a qualitative comparison measure that can be used to begin testing for convergent construct validity and external validity of quantitative indices. This rapid assessment allows researchers to critique how quantitative indices reflect individual communities, and perhaps predict their validity within a larger cluster of related communities. The method is an important first step in assessing the validity of quantitative indices.

As an illustration of this method, we have applied it to the quantitative indices developed by NMFS' social scientists for Alaska fishing communities. Identification of several inconsistent index components, both in terms of construct reliability (across researchers), construct validity (between quantitative and qualitative measures), and external validity (across communities), highlights the importance of further theoretical development of these particular elements of the well-being construct prior to their application in decision-making or trend analysis. Application of this methodology to indices and indicators developed in other research contexts and types of communities has the potential to highlight similar areas where improvement is needed.

This index validation method reveals instances in which quantitative indices may have been inadequate at describing local conditions related to vulnerability and resilience. For example, although results from 7 of the 13 communities exhibited poor or inconclusive external validity, it does not necessarily provide conclusive evidence that the method used in building the indices is inherently flawed. Communities are diverse and making generalizations on a macro scale is difficult. Context plays an important role in validity (Guhn et al., 2011), and a variable that adequately represents an index construct for one community or case may not be acceptable for another. The rapid assessment methodology outlined in this paper allows researchers to identify strengths and weaknesses within such indices themselves, and thus direct efforts towards uncovering why an index worked for one community, but not another.

Groundtruthing field methods thus serve multiple purposes. In addition to allowing for the development of a qualitative comparison measure that enables researchers to test for convergent construct validity, the qualitative data can provides meaning and context that can help researchers understand why certain index components failed to demonstrate sufficient construct and/or external validity. This index validation test affirms that it is not enough to simply create an index of community well-being, since that index requires place-specific meaning if it is to be used in explaining real-world phenomena or projecting community-based responses to SES-directed perturbations. Moreover, a detailed exploration of how qualitative constructs link broadly derived indices with more nuanced characteristics found in individual communities can assist in determining the usefulness of such indices as a management tool.

The results and discussion presented in this paper set the stage for a detailed content analysis that can inform additional theoretical development and refinement of all stages of NMFS' index development methods, including selection of variables, PCFA, cluster analysis, and qualitative comparison, as well as further assessment of construct validity, through detailed content analysis of qualitative interview data collected during fieldwork. To provide better context for interpretation of our quantitative indices, future work will include an intensive content analysis of transcripts and field notes collected in Step 3. The rapid assessment described in this paper will also support the process of content analysis through identifying constructs that were both contentious among the research team members, and/or poorly understood in terms of their relationship to the indices. Further, the results provide substantial evidence for the importance of groundtruthing quantitative indices in general so they may be better calibrated to reflect the communities or specific cases they seek to measure.

Acknowledgements

We would like to express our sincere gratitude to the residents in each of the communities we visited for their participation in this study. We also greatly appreciate for the help that Ben Fissel, Alan Haynie, the AFSC Publications Unit, and the reviewers and editorial staff of this journal provided in the development of this paper. This research has been fully funded by the NMFS Office of Science and Technology.

References

Adcock, R. and D. Collier (2001). Measurement validity: A shared standard for qualitative and quantitative research. In American Political Science Review 95(3): 529-546). Cambridge University Press.

ADF&G (Alaska Department of Fish and Game). (2011d). Community Subsistence Information System (CSIS). ADF&G Division of Subsistence. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. Accessed February 2011 from: http://www.adfg.alaska.gov/sb/CSIS/.

ADF&G (Alaska Department of Fish and Game) Alaska Department of Fish and Game. (2011a). Alaska sport fish and crew license holders, 2000 – 2010. ADF&G Division of Administrative Services. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

ADF&G (Alaska Department of Fish and Game) Alaska Department of Fish and Game. (2011b). Data on Alaska fish processors. ADF&G Division of Commercial Fisheries. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

ADF&G (Alaska Department of Fish and Game)Alaska Department of Fish and Game. (2011c). Alaska sport fish guide licenses and businesses, 2000 – 2010. ADF&G Division of Administrative Services. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

ADLWD (Alaska Department of Labor and Workforce Development). (2011). Alaska community population estimates, 2001 – 2009. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. http://labor.alaska.gov/research/pop/popest.htm (Accessed February and October 2011).

ADLWD (Alaska Department of Labor and Workforce Development). (n.d.). *Alaska Local and Regional Information Network.* Retrieved from http://live.laborstats.alaska.gov/alari/.

Alaska Commercial Fisheries Entry Commission. (2015). Alaska fish ticket data. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

Alaska Department of Community and Regional Affairs. (n.d.). *Community Database Online.* Retrieved from http://commerce.state.ak.us/cra/DCRAExternal.

Alaska Department of Education and Early Development. (n.d.). *Statistics and Reports.* Retrieved from http://eed.alaska.gov/stats/.

Alaska Fisheries Information Network. (n.d.). *AKFIN Answers.* Retrieved from www.akfin.org.

Allison, E. H., and Ellis, F. (2001). The livelihoods approach and management of small-scale fisheries. *Marine Policy* 25:377-388.

Allison, E. H., Perry, A. L., Badjeck, M. C., Adger, W. N., Brown, K., Conway, D., Halls, A., Pilling, G., Reynolds, J., Andrew, N., and Dulvy, N. K. (2009). Vulnerability of national economies to the impacts of climate change on fisheries. *Fish and Fisheries, 10,* 173-196.

Andrews, F. M., and Withey, S. B. (1976). *Social indicators of well-being – Americans' perception of quality of life*. Plenum Press. New York, NY.

Ban, N. C., Mills, M., Tam, J., Hicks, C., Klain, S., Stoeckl, N, Bottrill, M., Levine, J., Pressey, R. Satterfield, T. and Chan, K. (2013). A social-ecological approach to conservation planning: embedding social considerations. *Frontiers in Ecology and the Environment* 11, (4):194-202.

Berkes, F., and Jolly, D. (2001). Adapting to climate change: Social-ecological resilience in Canadian western arctic community. *Conservation Ecology* 5(2):18-32.

Biedenweg, K., Hanein, A., Nelson, K., Stiles, K., Wellman, K., Horowitz, J., and Vynne, S. (2014). Developing human wellbeing indicators in the Puget Sound: Focusing on the watershed scale. *Coastal Management* 42(4):374-390.

Boyd, H., and Charles, A. (2006). Creating community-based indicators to monitor sustainability of local fisheries. *Ocean and Coastal Management* 49:237-258.

Brookfield, K., Gray, T., and Hatchard, J. (2005). The concept of fisheries-dependent communities. A comparative analysis of four UK case studies: Shetland, Peterhead, North Shields and Lowestoft. *Fisheries Research* 72:55-69.

Carmines, E. G., and Zeller, R. A. (1981). *Reliability and validity assessment*. Safe. Beverly Hills, CA.

CFEC (Alaska Commercial Fisheries Entry Commission). (2011). *Alaska commercial fishing permits, permit holders, and vessel licenses, 2000 – 2010*, in: Data compiled by Alaska Fisheries Information Network for the Alaska Fisheries Science Center (Ed.), Seattle, WA.

Charles, A., Burbidge, C., Boyd, H., and Lavers, A. (2009). *Fisheries and the marine environment in Nova Scotia: Searching for sustainability and resilience.* GPI Atlantic. Halifax, Nova Scotia.

Charmaz, K. (2008). Constructionism and the Grounded Theory. In J. A. Holstein and J. F. Gubrium (Eds.), *Handbook of Constructionist Research* (397-412). New York: The Guilford Press.

Clay, P. M., and McGoodwin, J. R. (1995). Utilizing social sciences in fisheries management. *Aquatic Living Resources* 8(3):203-207.

Cloquell-Ballester, V., V Cloquell-Ballester, R. Monterde-Diaz and M. Santamarina-Siurana. 2006. Indicators validation for the improvement of environmental and social impact quantitative assessment. *Environmental Impact Assessment Review* 26(1): 79-105.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70:213–220.

Colburn, L. L., Abbott-Jamieson, S., and Clay, P. M. (2006). Anthropological applications in the management of federally managed fisheries: Context, institutional history, and prospectus. *Human Organization* 65(3):231-239.

Colburn, L. L., and Jepson, M. (2012). Social indicators of gentrification pressure in fishing communities: A context for social impact assessment. *Coastal Management* 40(3):289-300.

Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin* 88:322-328.

Connidis, I. (1984). The construct validity of the Life Satisfaction Index a and Affect Balance Scales: A serendipitous analysis. *Social Indicators Research* 15(2):117-129.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative and mixed methods approaches* (2nd ed.). Thousand Oaks: Sage Publications.

Creswell, J. W., Klassen, A. C., Plano-Clark, V. L., and Smith, K. C. (2011). *Best practices for mixed methods research in the health sciences.* Bethesda: National Institutes of Health.

Cutter, S., B. Boruff and W. L. Shirley (2003). Social vulnerability to environmental hazards. *Social Science Quarterly* 84(2):242-261.

ESRI. (2011). ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Fall, J.A. and D. Koster. (2011). Subsistence harvests of Pacific halibut in Alaska, 2009. Alaska Department of Fish and Game Division of Subsistence, Technical Paper No. 357, Anchorage. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle.

Fall, J.A., C. Brown, N. Braem, J.J. Simon, W.E. Simeone, D.L. Holen, L. Naves, L. Hutchinson-Scarborough, T. Lemons, and T.M. Krieg. (2011, revised). Alaska subsistence salmon fisheries 2008 annual report. Alaska Department of Fish and Game Division of Subsistence, Technical Paper No. 359, Anchorage. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle.

Fekete, A. 2009. Validation of a social vulnerability index in context to river-floods in German. *Natural Hazards and Earth System Science*, 9(2):393-403.

Fleiss, J. L., Levin, B., and Paik. M. C. (2003). *Statistical methods for rates and proportions*. (3rd ed.). New York: Wiley.

Frost, Kathy J., and Suydam, Robert S. (2010). Subsistence harvest of beluga or white whales (Delphinapterusleucas) in northern and western Alaska, 1987–2006. Journal of Cetacean Research and Management 11(3): 293–299.

Giddings, L. S., and Grant, B. M. (2007). A Trojan horse for positivism? A critique of mixed methods research. *Advances in Nursing Science* 30(1):52-60.

Glaser, B. (1992). *Basics of grounded theory analysis: Emergence vs. forcing*. Mill Valley, CA: Sociology Press.

Guest, G., Bunce, A., and Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18:59-82.

Guhn, M., Zumbo, B. D., Janus, M., and Hertzman, C. (2011). Validation theory and research for a population-level measure of children's development, well-being, and school readiness. *Social Indicators Research* 103(2):183-191.

Gunderson, L.H., and Holling, C.S. (Eds.). (2002). *Panarchy: Understanding Transformations in Human and Natural Systems*. Washington, D.C.: Island Press.

Healey, M. C. (2009). Resilient salmon, resilient fisheries for British Columbia, Canada. *Ecology and Society* 14(1):2.

Heath, H., and Cowley, S. (2004). Developing a grounded theory approach: a comparison of Glaser and Strauss. *International Journal of Nursing Studies* 41:141-150.

Himes-Cornell, A., and Kasperski, S. (2015). Assessing climate change vulnerability in Alaska's fishing communities. *Fisheries Research* 162:1-11.

Himes-Cornell, A., and Kasperski, S. (2016). Using socio-economic and fisheries involvement indices to better understand Alaska fishing community well-being. *Coastal Management* 44(1): 36-70.

Himes-Cornell, A., Hoelting, K., Maguire, C., Munger-Little, L., Lee, J., Fisk, J., Felthoven, R., Little, P. (2013). *Community profiles for North Pacific Fisheries – Alaska.* U.S. Department of Commerce, NOAA Tech. Memo. National Marine Fisheries Service-AFSC-259, Volumes 1-12.

Himes-Cornell, A. and K. Kent. (2014a). Involving Fishing Communities in Data Collection: A Summary and Description of the Alaska Community Survey, 2011. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-284, 171 p.

Himes-Cornell, A., K. Kent. (2014b). "Involving Fishing Communities in Data Collection: A Summary and Description of the Alaska Community Survey, 2010." U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-280, 170 p.

Imperial, M. T., and Yandle, T. (2005). Taking institutions seriously: Using the IAD framework to analyze fisheries policy. *Society and Natural Resources* 18:493-509.

Jacob, S., Weeks, P., Blount, B. G., and Jepson, M. (2010). Exploring fishing dependence in Gulf Coast communities. *Marine Policy* 34(6), 1307-1314.

Jacob, S., Weeks, P., Blount, B. G., and Jepson, M. (2013). Development and evaluation of social indicators of vulnerability and resiliency for fishing communities in the Gulf of Mexico. *Marine Policy* 37:86-95.

Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters* 31(8):651-666.

Jentoft, S. (2006). Beyond fisheries management: The Phronetic dimension. *Marine Policy* 30:671-680.

Jepson, M. and S. Jacob. (2007). Social indicators and measurements of vulnerability for Gulf Coast fishing communities. *NAPA Bulletin* 28:57-68.

Jepson, M., and Colburn, L. L. (2013). Development of social indicators of fishing community vulnerability and resilience in the US Southeast and Northeast regions. US Department of Commerce, NOAA Technical Memorandum NMFS-F/SPO-129, 64p.

Johnson, R. B., and Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher* 33(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. J., and Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research* 1(2):112-133.

Kelly, P. M., and Adger, W. N. (2000). Theory and practice in assessing vulnerability to climate change and facilitating adaptation. *Climatic Change* 47(4):325-352.

Landis, J. R., and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*:159-174.

Lowe, M. E. (2008). Crab rationalization and potential community impacts of vertical integration in Alaska's fisheries. In M. E. Lowe, and C. Carothers (Eds.), *Enclosing the fisheries: People, places, and power*. Symposium 68. Bethesda, MD: American Fisheries Society.

Lyubomirsky, S. and H. Lepper. 1999. A measure of subjective happiness: Preliminary reliability and construct validation. *Social indicators research*, 46.2: 137-155.

Magnuson-Stevens Fishery Conservation and Management Act of 2007. § 301, 16 U.S.C. § 1851 (2007).

Marshall, N. A., and Marshall, P. A. (2007). Conceptualization and operationalizing social resilience within commercial fisheries in northern Australia. *Ecology and Society* 12(1):1.

Martin, I. E. (2008). Resilience in Lower Columbia River Salmon Communities. *Ecology and Society* 13(2):23.

Mason, M. (2010). Sample size and saturation in PhD studies using qualitative interviews. *Forum: Qualitative Research* 11(3).

McGinn, T., Wyer, P. C., Newman, T. B., Keitz, S., and Leipzig, R. (2004). Tips for learners of evidence-based medicine: 3 Measures of observer variability (kappa statistic). *Canadian Medical Association Journal* 171(11):1369-1373.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica* 22(3):276-282.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research* 45:35-44.

Miller, F., Osbahr, H., Boyd, E., Thomalla, F., Bharwani, S., Ziervogel, G., Walker, B., Birkmann, J., van der Leeuw, S., Rockstrom, J., Hinkel, J., Downing, T., Folke, C., and Nelson, D. (2010). Resilience and vulnerability: Complementary or conflicting concepts? *Ecology and Society* 15(3):11.

Mills, J., Bonner, A., and Francis, K. (2006). The Development of constructivist grounded theory. *International Journal of Qualitative Methods* 5(1):25-35.

Mingers, J. (1980). Towards an appropriate social theory for applied systems thinking: critical theory and soft systems methodology. *Journal of Applied Systems Analysis* 7:41-50.

Morzaria-Luna, H. N., Turk-Boyer, P., and Moreno-Baez, M. (2013). Social indicators of vulnerability for fishing communities in the northern Gulf of California, Mexico: Implications for climate change. *Marine Policy* 45:182-193.

NMFS (National Marine Fisheries Service). (2011a). Alaska Individual Fishing Quota (IFQ) permit data. NMFS Alaska Regional Office. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

NMFS (National Marine Fisheries Service). (2011b). Data on Limited Liability Permits, Alaska Federal Processor Permits (FPP), Federal Fisheries Permits (FFP), and Permit holders. NMFS Alaska Regional Office. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

NMFS (National Marine Fisheries Service). (2011c). American Fisheries Act permit data. NMFS Alaska Regional Office. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

NMFS (National Marine Fisheries Service). (2011d). Catch Accounting System data. NMFS Alaska Regional Office. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle. [URL not publicly available as some information is confidential.]

NMFS (National Marine Fisheries Service). (2014). Fisheries of the Exclusive Economic Zone Off Alaska: Eligible communities, halibut IFQ regulatory area location, community governing body that recommends the CDQ, and the fishing programs and associated areas where a CDQ representing an eligible community may be permitted to participate. Updated December 8, 2014. 50 CFR Part 679: Table 21. Accessed July 2015 from: http://alaskafisheries.noaa.gov/rr/tables/tabl21.pdf.

Pollnac, R., Abbott-Jamieson, S., Smith, C., Miller, M.L., Clay, P.M., and B. Oles, B. (2006). Toward a Model for Fisheries Social Impact Assessment. *Marine Fisheries Review* 68:1-18.

Prell, C., Hubacek, K., and Reed, M. (2009). Stakeholder analysis and social network analysis in natural resource management. *Society and Natural Resources* 22(6):501-518.

Reed, M. S., Fraser, E. D., and Dougill, A. J. (2006). An adaptive learning process for developing and applying sustainability indicators with local communities. *Ecological Economics* 59(4):406-418.

Sepez, J., Norman, K., Poole, A., and Tilt, B. (2006). Fish scales: scale and method in social science research for North Pacific and west coast fishing communities. *Human Organization* 65(3);280-293.

Sherrieb, K., F. Norris and S. Galea. 2010. Measuring capacities for community resilience. *Social Indicators Research* 99: 227-247.

Smith, S. L., Pollnac, R. B., Colburn, L. L., and Olson, J. (2011). Classification of coastal communities reporting commercial fish landings in the US Northeast region: Developing and testing a methodology. *Marine Fisheries Review* 73(2):41-61.

StataCorp. (2011). *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP.

Tate, E. 2012. Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards* 63(2): 325-347.

Tate, E. 2013. Uncertainty analysis for a social vulnerability index. *Annals of the Association of American Geographers.* 103(3): 526-543.

U.S. Census Bureau. (2000). *Profile of Selected Social, Economic and Housing Characteristics for All Places in Alaska.* 2000 Decennial Census for all places in Alaska. Accessed August and October 2011: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.

U.S. Census Bureau. (2010a). *Profile of selected social, economic and housing characteristics for all places in Alaska.* 2005-2009 American Community Survey.

U.S. Census Bureau. (2010b). *Profile of Selected Social, Economic and Housing Characteristics for All Places in Alaska.* 2010 Decennial Census for all places in Alaska. Accessed August and October 2011: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.

USFWS (United States Fisheries and Wildlife Service). (2011). Marking, Tagging and Reporting Program data bases for northern sea otter, Pacific Walrus and polar bear. Office of Marine Mammals Management. Anchorage, Alaska. Data compiled by Alaska Fisheries Information Network for Alaska Fisheries Science Center, Seattle.

Viera, A. J., and Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine* 37(5):360-363.

Table 1. Topics Included for Each Interview Protocol Type.

Protocol	Interview topics
General (short form)	• Characterizing the community • Important issues facing the community • How community has changed over the past 5-10 years • How residents get along and deal with disagreements • Strengths and weaknesses of community • Future of the community
Commercial fishing	• How and where fish are off-loaded • Fishing supplies bought in and outside community • Relationship between fishermen in community • Changes seen in fishing historically vs. today • Places or occasions where commercial fishermen and/or their families gather • Location of local commercial fishermen's official residence
Recreational fishing (charters and private anglers)	• Description of charter fishing clientele, crewmembers • Relationship between fishermen in community • How catch is used and who it is shared with • Fishing supplies bought in and outside community • Travel needed to purchase supplies • Changes seen in recreational fishing historically vs. today • Importance of recreational fishing to culture of community
Subsistence fishing	• Species caught for subsistence locally • Informant role/experience in subsistence fishing • How catch is used and who it is shared with • Distance to fishing grounds • Reason for undertaking subsistence fishing • Places or occasions where subsistence fishermen and/or their families gather • Changes seen in recreational fishing historically vs. today
Local business	• Goods and services provided or get from local fishermen
City leadership	• Important sources of jobs and income in community • Importance of fishing for the economy and culture of community • Major community fishing-related events • Comparison of current fishing industry compared to historical fishing • Policies in place (at any level of government) to encourage or restrain the fishing industry • Role of climate change and fishing in the community's comprehensive plan • Expected effects of climate change on community

Table 2. Total Number of Interviews Conducted across Interview Protocols and Communities.

Protocol / *Community*	General (short form)	Commercial fishing	Recreational fishing	Subsistence fishing	Business operation	Community leader	Total interviewed
Aleknagik	*11*	*5*	*3*	*6*	*0*	*3*	13
Dillingham	*35*	*12*	*4*	*13*	*9*	*8*	40
Kenai	*13*	*3*	*0*	*1*	*6*	*2*	15
King Salmon	*14*	*3*	*8*	*3*	*4*	*3*	14
Kodiak	*44*	*14*	*2*	*2*	*9*	*5*	49
Naknek	*23*	*10*	*2*	*8*	*4*	*5*	24
Ouzinkie	*15*	*6*	*1*	*6*	*0*	*2*	18
Port Graham	*5*	*1*	*2*	*4*	*1*	*2*	10
Port Lions	*15*	*6*	*6*	*4*	*0*	*2*	19
Sand Point	*23*	*15*	*1*	*7*	*4*	*5*	27
Seldovia	*22*	*6*	*5*	*2*	*1*	*2*	26
Soldotna	*15*	*2*	*6*	*0*	*5*	*1*	16
South Naknek	*12*	*8*	*1*	*6*	*1*	*4*	15
Total protocols administered	*247*	*91*	*41*	*62*	*44*	*44*	*286 indiv.* / *529 protocols*

*Table 3. Kappa Statistic Interpretation Scale (*Landis and Koch 1977).

Kappa Statistic	**Agreement**
< 0	Less than chance agreement
0.01 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 0.99	Almost perfect agreement

Table 4. Social Vulnerability Principal Components Factor Analysis (Armor's Theta = 0.959).

Component Constructs	**Five Highest Loading Variables**	**Eigenvalue**	**% variation explained**	**Cum. % variation explained**
Community Size	Total employment Peak quarterly # of workers Population Total households # of workers employed in all four quarters	15.88	20%	20%
Infrastructure	Clinic present Water services Sewer services Post office present Piped water utilities	8.87	11%	32%
Rural/Village Character	Avg. household size (2005-2009 ACS) Avg. household size (2000 Census) % Population under 18 Alcohol control laws % Speaking primary language other than English	7.56	9%	41%
Poverty	% Living below poverty line (per capita) % Families living below poverty line % Households earning under $10k % Unemployed % Occupied households lacking plumbing	7.17	9%	50%
Transient Population	% Living in another country one-year prior % Living in another state one-year prior % Population black or African American % of households renting % Living in same house one-year prior	3.30	4%	54%
Foreign Born Asian Population	% Foreign born population % Population Asian	3.24	4%	59%
Retirees/Low Female Workforce	% Households with 65 or older resident % Receiving social security % 25 and older with less than 9th grade education % Retired % Employed females 16 and over	3.04	4%	62%

Table 5. Fishery involvement principal components factor analysis (Armor's Theta = 0.975).

Component Constructs	Five Highest Loading Variables	Eigenvalue	% variation explained	Cum. % variation explained
Fishery Participation (total)	Vessels homeported Vessels owned by residents Crew licenses Total CFEC permits fished Total CFEC permit holders	15.91	22%	22%
Fishery Participation (per capita)	FFP permit holders Sablefish IFQ account holders Vessels owned by residents Vessels homeported Halibut IFQ account holders	11.27	15%	37%
Crab, AFA, and FPP	Crab permits fished Crab permits held by residents Crab IPQ account holders AFA permit holders (per capita) AFA permits fished (per capita)	8.38	11%	49%
Sportfishing (per capita)	Sport fish licenses sold Sport fish licenses held	3.80	5%	54%
FPP (per capita) and Seat Otter Subsistence (per capita)	FPP permits used FPP permit holders # of sea otters harvested	3.73	5%	59%
Landings (per capita), Vessels (per capita), and Processors (per capita)	Vessels making landings # of shoreside processors receiving landings Total net pounds landed Total ex-vessel value of landings	3.43	5%	64%
Marine Mammal (per capita) and Salmon Subsistence (per capita)	Marine mammals harvested Subsistence salmon permits returned Marine mammal pounds harvested # of subsistence salmon harvested	3.34	5%	68%
Federal Crab Permits (per capita) and Beluga Subsistence (per capita)	Crab permits fished Crab permit holders Subsistence beluga harvested	2.85	4%	72%

Note: If "per capita" is listed next to a construct in column 1, assume all variables related to that construct are measured as such; otherwise, individual per capita variables will be listed as such in column 2.

Table 6. Results of the construct reliability test.

	P < 0.05*	Average kappa	Rank
Social construct			
Community Size	Yes	0.42	Moderate Agreement
Infrastructure	Yes	0.52	Moderate Agreement
Rural/Village Character	Yes	0.74	Substantial Agreement
Poverty	Yes	0.48	Moderate Agreement
Transient Population	Yes	0.31	Fair Agreement
Foreign Born Asian Population	Yes	0.55	Moderate Agreement
Retirees	Yes	0.22	Fair Agreement
Low Female Workforce	No	-0.04	Inconclusive
Fisheries involvement construct			
Fishery Participation	Yes	0.52	Moderate Agreement
Crab, AFA, and FPP	Yes	0.42	Moderate Agreement
Sportfishing	Yes	0.37	Fair Agreement
Processor Activity	Yes	0.62	Substantial Agreement
Sea Otter Harvesting	Yes	0.26	Fair Agreement
Perceived Amount of Landings	Yes	0.75	Substantial Agreement
Vessels Located in Community	Yes	0.51	Moderate Agreement
Marine Mammal Harvesting	Yes	0.24	Fair Agreement
Salmon Subsistence	No	-0.05	Inconclusive
Number of Crab Permits	Yes	0.36	Fair Agreement
Beluga Harvesting	Yes	0.19	Slight Agreement

* P-values were not averaged. If at least one test produced a statistically significant result of $P < 0.05$, then the corresponding kappa was accepted.

Table 7. Results of Inter-Observer Reliability and Convergent Construct Validity Tests.

Inter-observer reliability test			***Construct validity test***		
Community	**Average Kappa**	**P < 0.05***	**Average Kappa**	**P < 0.05***	**Result**
South Naknek	0.5959	Yes	0.11	Yes	Poor Construct Validity
Soldotna	0.5056	Yes	0.44	Yes	Moderate Construct Validity
Seldovia	0.2083	No	-0.20	Yes	Inconclusive
Sand Point	0.3638	Yes	0.41	Yes	Moderate Construct Validity
Port Lions	0.3982	Yes	0.11	No	Poor Construct Validity
Port Graham	0.7121	Yes	0.34	Yes	Fair Construct Validity
Ouzinkie	0.5552	Yes	0.21	No	Poor Construct Validity
Naknek	0.2294	Yes	0.15	No	Poor Construct Validity
Kodiak	0.6154	Yes	0.06	No	Poor Construct Validity
King Salmon	0.4526	Yes	0.37	Yes	Fair Construct Validity
Kenai	0.2091	Yes	0.32	Yes	Fair Construct Validity
Dillingham	0.0796	Yes	0.06	No	Inconclusive
Aleknagik	0.5291	Yes	0.36	Yes	Fair Construct Validity

* P-values were not averaged. If at least one test produced a statistically significant result of $P < 0.05$, then the corresponding kappa was accepted.