Field Methods

http://fmx.sagepub.com/

Competence, Agreement, and Luck: Testing Whether Some People Agree More with a Cultural Truth than Do Others

Daniel J. Hruschka and Jonathan N. Maupin

Field Methods published online 19 September 2012

DOI: 10.1177/1525822X12453514

Field Methods
00(0) 1-17

DOI: 10.1177/1525822X12453514
http://fm.sagepub.com
SAGE

Competence, Agreement, and Luck: Testing Whether Some People Agree More with a Cultural Truth than Do Others

Daniel J. Hruschka[1] and Jonathan N. Maupin[1]

Abstract

The cultural consensus model (CCM) is a frequently used model of cultural diversity, which predicts how individuals from a common cultural background would share knowledge in a specific domain. Cultural competence, or the degree to which an individual agrees with a local cultural truth, is a central concept in CCM, and many uses of competence estimates rest on the assumption that they reflect real variation among individuals in their knowledge of a cultural truth. However, Weller has shown that even in situations where there are no real differences among individuals in their underlying competences, the CCM will still estimate individual competences that appear to vary, sometimes substantially. To address this issue, we describe a test of the null hypothesis that there is no real difference

[1] School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA

Corresponding Author:
Daniel J. Hruschka, School of Human Evolution and Social Change, Arizona State University, PO Box 872402, Tempe, AZ 85287, USA
Email: dhruschk@asu.edu

in competence among individuals. We also present tables with specific cut-offs for this test across a range of data set characteristics. These can help researchers decide whether a specific set of data warrants further analyses of individual differences in competence.

Keywords
competence, cultural consensus model, cultural consensus theory, cultural consensus analysis, test theory without an answer key

Introduction

Cultural consensus theory was developed in the mid-1980s to handle a recurring challenge in anthropology: to describe shared patterns of thought and behavior in a society while also capturing diversity in how people think and behave (Batchelder and Romney 1988; Romney et al. 1986). The theory relies on four basic assumptions about how people from a common cultural background share knowledge. First, it assumes that there is a single body of knowledge from which all people draw information, although no single person likely knows everything perfectly. Second, it assumes that differences in people's responses to questions about this body of knowledge arise from random guessing when they do not know the answer. Third, it assumes that each person has some fixed ability, or "competence," to correctly answer questions about this body of knowledge. Finally, the model assumes that there is no systematic collusion or influence among people as they respond to questions about the body of knowledge. That is, each person draws independently from the same body of knowledge.

With these assumptions and some further specification of how people respond to questions about a particular cultural domain, the theory can be written as a statistical model that makes predictions about how knowledge is distributed among a group of people. It is important to note that these assumptions permit a great deal of leeway in how we specify the exact mathematical model. At a very basic level, the original model was developed to deal with yes/no answers, multiple choice responses, and fill in the blank, but recent elaborations permit ranking and rating data (Batchelder et al. 2010).

A key concept in the cultural consensus model (CCM) is "cultural competence," or the probability that someone will know the correct answer to a question about the specific cultural domain. In brief, a person with a higher competence is more likely to answer correctly about the domain than someone with a lower competence. When applied to people's real responses about a body of data, the CCM produces an estimate of this value for each respondent in the study.

These estimated competence values have commonly been used in four ways. First, when the model estimates the culturally true answer key, it gives greater weight to the responses of those people who have larger competence estimates (Romney et al. 1986). Second, researchers often interpret the competence estimates as differential knowledge in the domain and then correlate these with individual-level variables such as age, education, or success in the domain (Romney 1999). Third, researchers have used competence scores to identify cultural informants with greater knowledge of a domain (Johnson 1990). A fourth, less common, use of competence estimates is to extract what is called residual agreement among respondents calculated by subtracting the predicted agreement (the product of informants' competence scores) from the observed agreement (Ross 2004).

These uses of cultural competence estimates rest on the assumptions that if people are assigned different estimates by the model, these somehow reflect real differences in people's knowledge of a particular domain. It is important here to distinguish between *estimated* competence and *actual* competence. In most practical applications of the CCM, researchers never deal with actual competence. Rather, they acquire estimates of individual competence from the model fit to their data. These estimates are presumed to reflect the actual ability, or competence, of individual respondents to provide the culturally correct answer. However, these estimates also reflect other things as well, most notably differences in the degree to which people randomly guessed correctly. As an example, consider a test involving 20 yes/no questions about flora in Madagascar (for which we already know the correct answer). If we ask these questions of 1,000 people who know nothing about flora in Madagascar, then some people will get 14 correct, while others will get 7 correct, even though there is no difference in their underlying abilities. These differences result purely from differential luck, and we would never consider interpreting these differences as differential expertise in plants from Madagascar. This extreme case illustrates that even when there is no knowledge underlying people's responses, some will perform better than others on a test due to random luck. In more familiar domains where everybody holds some knowledge, performance will fold in both this differential luck and any differential ability that may exist.

Given a specific set of data, it is entirely possible that any differences in estimated competence are due largely to differential luck. Indeed, Weller (1987) showed that even if everyone has exactly the same real competence in a domain, the CCM will produce *estimates* of competence that can vary a great deal between people. In short, the model will estimate a diversity of competence values for individuals, even when there are no real differences between respondents. In such cases, the estimated differences are most appropriately attributed to differential luck than to any real difference between people.

This finding poses a problem for investigators who would like to use differences in competence *estimates* as an indicator of real differences in knowledge. If the estimated differences do not actually reflect real differences, then it is not clear what added information one can glean about the answer key by giving greater weight to individuals with higher competence estimates. Moreover, if the estimated differences in competence only reflect random error, then it is not clear why one would expect correlations with other individual-level variables such as age or sex.

Testing Whether There Are Differences in Real Competences

In this article, we describe a simple model check to determine whether the individual competences estimated from a specific data set warrant further use or interpretation. The approach is based on earlier simulation approaches described by Weller (1987) and Brewer (1995) but provides specific cutoffs for varying confidence levels.

For the model check, we start by asking a simple question. If we have data from a group of individuals all of whom have the same real competence, what would the CCM tell us about their competences? What would it estimate? In the vast majority of cases, the CCM will estimate some differences between individuals, even when there is no real difference. Our question, then, concerns a matter of degree. If the CCM always estimates some variation in competence scores, how much variation in estimated competences would we need to see to be certain that there was really variation in ability among individuals?

Before describing the test, it is useful to outline how the consensus model makes predictions about people's response to yes/no questions about a specific cultural domain. The model first assumes that there is a set of questions that an investigator will ask about the domain. We will assume there are k questions. Second, there is a sample of n people who respond to each of these questions. Each person has a different probability of knowing the answer to a question. Let us call D_{Ann} the probability for the person named Ann. These D values are also known as *competences*. Furthermore, most implementations of the model assume that when people do not know the answer to a question that they guess "yes" half of the time.

Based on these assumptions, the model predicts that the respondent named Ann will respond "yes" to a specific question with the following probabilities:

1. If the correct answer to a specific question is "yes," then the probability of Ann saying "yes" to that question is $D_{Ann} + (1 - D_{Ann})0.5$. This means that Ann will know the correct answer with some probability, D_{Ann}. When she does not know the correct answer (with probability $1 - D_{Ann}$), she will guess "yes" half of the time.
2. If the correct answer is "no" to a question, then the probability of Ann saying "yes" to that question is $(1 - D_{Ann})0.5$. This means that Ann will only say yes if she was guessing, and in those cases, she will guess "yes" half the time.

Based on this model, if we specify two things—a set of correct answers and an individual competence for each person in a group—then we can generate the kinds of responses that we would expect from those people to those questions. Since the model generates responses probabilistically, there is no single set of responses generated from the model. Rather, we can generate a whole range of typical data sets for the model.

By generating a large number of typical data sets in this way, we can test whether an observed set of data is "atypical" for the specific set of assumptions we make (Brewer 1995; Weller 1987). For example, consider the Weller data set on contagiousness of diseases among 24 Guatemalan respondents (Romney 1999; Weller 1984). The respondents were asked 27 yes/no questions, and the average estimated competence in the group was 0.82.

Now let us generate a number of data sets with these same properties but assume everyone has the same mean competence of 0.82. When we make all competences equal, the only factor creating differences in individual responses is differential luck, or more specifically differences as a result of guessing. For each of these generated data sets, we can fit the traditional CCM and calculate how much variation there is in *estimated* competences (using the standard deviation of these estimated competences). In short, this approach tells us how much variation in competences the CCM estimates when there really is no variation in people's ability.

If we repeat this process thousands of times, we can see how frequently the CCM estimates different amounts of variation in competence for

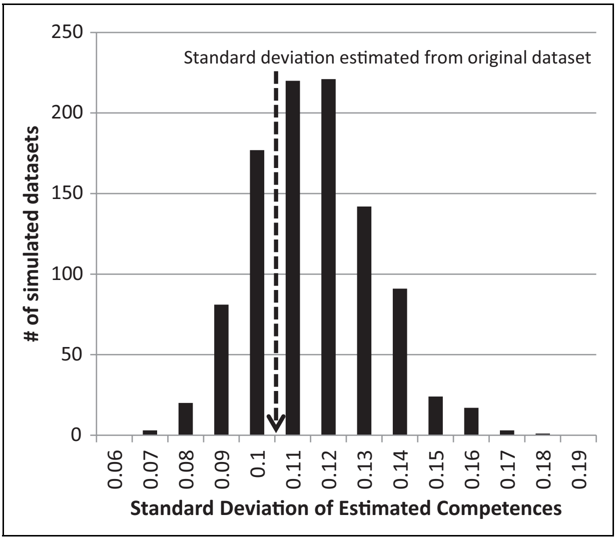

Figure 1. Histogram of standard deviations for competences estimated by the cultural consensus model when there is no real variation in competence. One thousand data sets generated by model with 27 yes/no questions, 24 respondents with competence = 0.82, and a true answer key with 14 "yes" responses.

respondents that really do not vary in competence. Figure 1 illustrates such a graph for 24 people responding to 27 yes/no questions, when the average competence is 0.82.

Figure 1 shows that even when respondents in a sample have no real difference in competence, the CCM will frequently estimate competences that show substantial variation among individuals in the group. In none of the 1,000 data sets did the CCM estimate that everyone had the same competence (e.g., standard deviation = 0). In about 50% of all cases, the standard deviation of estimated competences is greater than 0.10. This is near the top of the hump in the histogram in Figure 1.

Now, let us return to the Weller 1984 data set. When we fit the consensus model to Weller's data, it provides a competence estimate for each person, and the standard deviation of these estimates is 0.10. When we compare this value with the distribution in Figure 1, we see that this value is indeed quite typical if we assume that everyone has equal competences. Thus, Weller's Guatemala data are consistent with an assumption that everyone has equal competences.

We can refine this comparison by asking how big the observed variation in Weller's data set would need to be so that it is *not* typical of a group with equal competences. For example, how big would the observed variation in Weller's data set need to be to surpass some large number (let us say 95%) of data sets generated with an assumption of equal competence? For the distribution in Figure 1, this 95% cutoff would be found at a standard deviation of 0.13. This cutoff provides a test of whether the Weller data are typical of a data set where everyone has equal competence. If the observed standard deviation among competences in the Weller data set was larger than 0.13, then we could reject the assumption of equal competence with a good degree of confidence. In that case, it would be very unlikely to have observed such large variation in estimated competences if the real competences were all the same. On the other hand, if the observed standard deviation in competence scores is less than 0.13, then the Weller data set is not strikingly different from other data sets where all people have the same competence. In that case, there is little reason to believe that individuals in the Weller 1984 data set differ in real competence.

The observed standard deviation for the Weller 1984 data set is 0.104, which is below 0.13 and well within the typical range for comparable groups where people have the same real competence. Thus, the analysis did not reveal significant variability in competence. In such a case, we propose that investigators not use or analyze individual competence estimates. On the other hand, if the variation in competence estimates *had been* atypically large, then we could reject the assumption of no difference in competence among individuals. In such a case and in the absence of other indications that the competence estimates are not valid, the investigators can interpret individual differences in competence estimates as reflecting some real differences in competence.

Tables and an Algorithm

The cutoff described in the previous section depends on a number of factors, including the number of people in the data set, the number of questions asked, and the average competence. It also depends on a factor we only briefly mentioned in the earlier section—the rate at which individuals guess "yes" if they do not know the answer.

Table 1 lists the 95% cutoff for a range of values when we assume that everyone guesses "yes" half the time. This is the assumption in most factor analytic estimates of the model and what one gets from UCINET software package (Borgatti et al. 2002). Not surprisingly, the number of questions is

Table 1. Estimated 95% Cutoff for Standard Deviation of Competence Scores Should Be Based on the Number of Respondents (*N*), Number of Questions (*k*) and Average Competence of Individuals (Guessing Parameter $g = 0.50$, 50% Yes Answers in Answer Key, and 1,000 Simulated Data Sets)

N	*k*	Mean competence			
		0.5	0.6	0.7	0.8
25	25	0.23	0.21	0.18	0.15
25	50	0.16	0.14	0.13	0.11
25	100	0.11	0.10	0.09	0.07
25	200	0.08	0.07	0.06	0.05
50	25	0.20	0.18	0.16	0.14
50	50	0.15	0.13	0.12	0.10
50	100	0.10	0.09	0.08	0.07
50	200	0.07	0.07	0.06	0.05

Table 2. Estimated 95% Cutoff for Standard Deviation of Competence Scores Should Be Based on the Number of Respondents (*N*), Number of Questions (*k*) and Average Competence of Individuals (Guessing Parameter in Uniform Distribution 0.00 to 1.00, 50% Yes Answers in Answer Key, 1,000 Simulated Data Sets)

N	*k*	Mean competence			
		0.5	0.6	0.7	0.8
25	25	0.26	0.21	0.18	0.14
25	50	0.19	0.14	0.13	0.11
25	100	0.15	0.10	0.09	0.07
25	200	0.13	0.07	0.06	0.05
50	25	0.20	0.18	0.16	0.14
50	50	0.16	0.13	0.12	0.10
50	100	0.12	0.09	0.08	0.07
50	200	0.11	0.07	0.06	0.05

the most important factor for decreasing the standard deviation of estimated competences due to random error. As the number of questions increases, the random variation in estimated competences gets closer and closer to zero. However, the average competence as well as the number of participants has some influence on the cutoff as well.

Table 2 lists the 95% cutoffs if we assume that every person has a different probability of guessing "yes" that can range from 0 to 1. This cutoff may be more appropriate when it is not known whether there is individual variation in guessing. Indeed, until diagnostics are developed for estimating the range of individual guessing, we recommend using the more conservative Table 2. When mean competence is sufficiently high and the number of questions is sufficiently large, then the cutoffs in Table 2 are very close to those in Table 1. However, in cases of low competence, the cutoff is sensitive to the assumptions we make about individual variation in guessing.

To use the tables for your specific data set, first calculate the mean and standard deviation of competence scores when you run the consensus model on your data. Then identify the cutoff in the table for the values of n, k, and mean competence that most closely approximate your data set. If the standard deviation for your competence estimates is greater than the value in the entry, then you can reject the null model that people have uniform competence. However, if the standard deviation for your competence estimates is less than the value in the entry, then you *cannot* reject the null model that people have uniform competence. If the values are close, then one can interpret this the same way as one would a marginally significant p value of around .05.

Table 3 shows the test applied to seven data sets: Guatemalan women responding to questions about the contagiousness of diseases (Weller 1984); Mexican women responding to questions about signs, causes, consequences, and treatments for local illnesses (Garro 2000); Ojibway adults responding to questions about signs, causes, consequences, and treatments for diabetes (Garro 2000); and the knowledge of signs, causes, and treatments for postpartum hemorrhage among biomedically trained birth attendants (SBA), traditional birth attendants (TBA), and laywomen in Bangladesh (Hruschka et al. 2008).

In data sets with more than 50 questions, the model estimates competences with greater variation than one would expect, given an assumption of equal competences. This occurs for a range of mean competence levels. However, for data sets with 18 or 27 questions, there was not sufficient variation in estimated competences to reject the hypothesis of equal competence.

At the end of the article, we provide code for the R statistical package that will permit estimating the cutoff for a data set when we know: the number of respondents (N), the number of questions (k), the average competence (mean D), the number of yes responses in answer key (y), and the distribution of the guessing parameter (g). Investigators who find that their estimate for the true answer key diverges dramatically from 50% yes responses can use the available R code to recalculate the cutoff for the estimated number of yes responses (R Development Core Team 2011).

Table 3. The Test Applied to Seven Data Sets. Guatemalan Women's Knowledge of Contagiousness of Diseases (Weller 1984); Mexican Women Responding to Questions about Signs, Causes, Consequences, and Treatments for Local Illnesses (Garro 2000); Ojibway Adults Responding to Questions about Signs, Causes, Consequences, and Treatments for Diabetes (Garro 2000); Knowledge of Signs, Causes, and Treatments for Postpartum Hemorrhage Should Be among Medically Trained Birth Attendants (SBA); Traditional Birth Attendants (TBA); and Laywomen in Bangladesh (Hruschka et al. 2008)

Data set	*N*	*k*	*Y*	Mean competence	Std. dev. competence	95% cutoff (variable g)	Reject uniform competence?
Weller 1984	24	27	14	0.82	0.10	0.13	No
Garro 2000, Mexican women	20	396	198*	0.68	0.12	0.05	Yes
Garro 2000, Ojibway all statements	34	68	34*	0.58	0.16	0.12	Yes
Garro 2000, Ojibway causal statements	34	18	6	0.67	0.11	0.21	No
Hruschka et al. 2008, SBAs	14	234	98	0.74	0.09	0.07	Yes
Hruschka et al. 2008, TBAs	37	234	98	0.62	0.12	0.06	Yes
Hruschka et al. 2008, laywomen	98	234	108	0.58	0.16	0.07	Yes

Note: N = number of respondents; K = number of questions; Y = number of yes responses.[a] No information provided in publication. Assumed half responses are yes.

Recommendations

The main purpose of the proposed test is to determine when there is sufficient variation in estimated competence to reject the notion that real competences are all the same. When a data set does not have sufficiently large variation in estimated competences, investigators should be cautious about interpreting individual estimated competences as indicating differences between people.

If there are no real differences between individuals, there are several implications for data analysis. First, creating a weighted estimate of the answer key based on differential competence is an unnecessary complication, and one could get comparable results from taking the simple average of individual responses. Second, inferences about group differences in competence, especially those with borderline significance (e.g., p around .05) and resulting from several tests, should be interpreted with caution. Third, there is little evidence that such scores provide a basis for identifying "knowledgeable" key informants. Fourth, estimating residual agreement by partialing out an agreement matrix calculated from individual competences is an unnecessary complication. Indeed, using competence scores based largely on noise might add random error to the residual matrix.

If the test shows that there is sufficient variation in competence estimates to assume some individual differences, then it should be possible for researchers to examine competence scores as some measure of individual difference. However, it is important to keep in mind that there are other potential reasons why competence estimates may not mean what we think they do, including the fact that the entire model is misspecified. This article does not touch on any of these reasons. We hope that future work will focus on telling us when and how competence scores actually tell us something about meaningful individual differences in knowledge.

There are a number of things this test does not tell us. First, it does not tell us whether there is a single answer key (Hruschka et al. 2008). Rather, it asks whether there are individual differences in people's knowledge of an answer key. Second, it does not validate or invalidate mean competence for the population. It is still possible to use overall average competence scores as an estimate of how much individuals agree in a population. Indeed, a report describing the mean competence, the standard deviation in competence, the overall probability of responding yes, and the proportion of individuals responding yes to each of the individual questions are all useful statistics for a reader hoping to interpret the output of the model.

Finally, when we look at changes in 95% cutoffs by the number of respondents and number of questions, it is clear that increasing the number of questions is by far the most effective way to increase one's ability to identify individual differences in competence. This makes sense from a psychometric perspective. Each question is one probe for identifying individual differences and the more probes one uses, the more likely one will be able to distinguish between individuals. However, there are also important limits on the number of questions, most notably the burden on respondents and interviewers. Another assumption of current implementations of the model is that questions be similar in difficulty and adding more questions of comparable difficulty may not be possible. Thus, before simply adding more questions, a researcher should think about two important questions: (1) to what degree is identifying individual differences in competence an important part of the research design; and (2) if it is an important element, what does one expect the estimated mean competence and variation in competence to be? For the second question, the researcher may have to make an educated guess based on prior research. With these estimates, the researcher can identify the number of questions one would need to ask to ensure that one would reject an assumption of equal competences.

The simulations also indicate that increasing the number of respondents plays a much smaller role in reducing the random error in competence scores than does increasing the number of questions. This also makes sense from a psychometric perspective, where increasing the number of individuals will not dramatically increase the ability to discriminate between them.

Conclusion

Expanding on observations by Weller (1987) and Brewer (1995), this article shows that variation in estimated competences is not a clear indication that there are true differences in competence among individuals in a sample. Indeed, a population of individuals who all have the same actual competence can easily *look* like they have different competences based on the estimates from the model.

This article provides one test of whether individuals in a population have equal competences and thus whether it is worthwhile using estimated competences for other purposes such as weighting answer keys in the CCM, analyzing between-individual differences in competence, or using competence estimates for estimating residual agreement. Of course, this is only one of numerous checks, each of which can provide further insight into the model and the data. For example, Weller (2007) proposes a test for the existence of reliable between-individual variation in competence scores. For the test, split the items in half and run the CCM independently on each set of items. If there is reliable between-individual variation in competence, then individual competence estimates from the first run should correlate with those in the second run.

An examination of how cutoffs vary with different model parameters also raises two important points. First, the ability to estimate individual differences in competence relies most heavily on the number of questions asked. Depending on the average competence, one may need 100 or more questions to overcome sampling error. That said, for the studies we examined here, once there are more than 30–50 questions, it was possible to reject an assumption of equal competence. Of course, one's choice of the number of question depends on the specific goals of the research and whether identifying individual differences in competence is an important part of the research design.

Second, for certain combinations of low average competence and a large numbers of questions, the test will depend on one's assumption about how people guess. Current implementations of the CCM make the assumption that people guess "yes" at a uniform rate. Future work that provides insight into how much rates of guessing vary will help inform the best approach to choosing cutoffs.

Finally, we have described a test for the traditional formal process model for yes/no responses. Recent work on formal process models for other data types, such as ranking, similarity judgments, matching, and continuous responses, will permit extending this test to other data types in the future (Batchelder et al. 2010).

Appendix A

R Code for Estimating 95% Cutoff

```
##Definition of function that calculates correlation matrix for response data randomly generated based on
##n individuals with uniform competence (averaged) answering k yes/no questions on an
## answer key with y yes responses. Ming is minimum value for uniform distribution for guessing parameter
##Maxg is maximum value for uniform distribution for guessing parameter.
  CreateMat<-function(n,k,y,averaged,ming,maxg){
   Key<- rep(c(1,0),c(y,k-y))
   Competence<-rep(averaged, times = n)
```

```
  Guessing<-runif(n, ming, maxg)
  ## create response vector
  Response <- matrix(0,n,k)
  for (i in 1:n)
    {
      for (j in 1:k)
      {
        if (Key[j] == 1) probYes <- Competence[i]+(1-Competence[i])*
Guessing[i]
        else probYes <- (1-Competence[i])*Guessing[i]
        Threshold <- runif(1,0,1)
        if (Threshold <= probYes) Response[i,j] <- 1
        else Response[i,j] <- 0
      }
    }
  ## create matrix based on matching adjusted for chance agreement
  CorrMatrix <- matrix(0,n,n)
  for (i in 1:n)
  {
    for (j in 1:n)
    {
      N00 <- 0
      N10 <- 0
      N01 <- 0
      N11 <- 0
      for (m in 1:k)
      {
        if (Response[i,m] == 0 && Response[j,m] == 0) N00 <-
N00+1
        if (Response[i,m] == 0 && Response[j,m] == 1) N01 <-
N01+1
        if (Response[i,m] == 1 && Response[j,m] == 0) N10 <-
N10+1
        if (Response[i,m] == 1 && Response[j,m] == 1) N11 <-
N11+1
      }
      denom <- (y/k)*(1-y/k)*k*(k-1)
      CorrMatrix[i,j] <- (N00*N11-N01*N10)/denom

    }
```

```
}
return(CorrMatrix)
}
##Definition of minres procedure
MinRes <-function(CorrMatrix,n) {
 avect <- runif(n, 0, 1)
   maxdelta <- 1
   while(0.01 < maxdelta) {
      maxdelta <- 0
      for (i in 1:n)    {
         pastterm <- avect[i]
         newnumer <-0
         newdenom <- 0
         for (j in 1:n) {
            if (j != i) newdenom <- newdenom+avect[j]*avect[j]
            if (j != i) newnumer <- newnumer+avect[j]*CorrMatrix[i,j]
         }
         avect[i] <- newnumer/newdenom
         if (avect[i] > 1) avect[i] <- 1
         if (avect[i] < -1) avect[i] <- -1
         maxdelta <- maxdelta+ abs(avect[i]-pastterm)
      }
   }
   return(avect)
}
##Residual of matrix
residual <-function(CorrMatrix,avect,n){
   residual <- 0
   for(i in 1:n){
      for (j in 1:n){
         error <- (CorrMatrix[i,j]-avect[i]*avect[j])
         if (i != j) residual <- residual + error*error
      }
   }
   return(residual)
}
##Definition of function that estimates Standard Deviation of Competences for a randomly generated
##dataset based on n individuals with uniform competence (averaged) answering k yes/no questions on an ## answer key with y yes
```

responses. Ming is minimum value for uniform distribution for guessing parameter

```
##Maxg is maximum value for uniform distribution for guessing parameter.
SDest <- function(n,k,y,averaged,ming,maxg){
   CorrMatrix <- CreateMat(n,k,y,averaged,ming,maxg)
   ## perform factor analysis
 ##uls <- fa(CorrMatrix,nfactors = 1,fm="minres")
   uls <- MinRes(CorrMatrix,n)
   ## return standard deviation of first factor scores
STDest <- sd(uls)
return(STDest)
 }
##Definition of function that generates "iternum" datasets with uniform individual competences
##and records standard deviations for competence estimates for each of the datasets
##the function returns a vector the standard deviations for the entire collection of datasets.
SDDist <- function(n,k,y,averaged,ming,maxg,iternum){
SDlist <- rep(0, times = iternum)
   for (i in 1:iternum)
   {
      SDlist[i] <-SDest(n,k,y,averaged,ming,maxg)
   }
   return(SDlist)
}
##This code calls the above function for datasets having properties of the Weller Guatemala data (1984)
Standard_Deviation <- SDDist(24,27,14,0.82,0,1,1000)
##This code takes the output vector from the previous line and shows the 95% cutoff (under 95%).
##It also shows all 5%-iles for the distribution
quantile(Standard_Deviation,probs=seq(0,1,0.05))
```

Acknowledgment

We thank Amber Wutich, Meredith Gartin, and Ashlan Falletta-Cowden for helpful comments on earlier drafts.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Hruschka gratefully acknowledges support from NSF grant BCS-1150813.

References

Batchelder, W. H., and A. K. Romney. 1988. Test theory without an answer key. *Psychometrika* 53:71–92.

Batchelder, W. H., A. Strashny, and A. K. Romney. 2010. Cultural consensus theory: Aggregating continuous reponses in a finite interval. *Lecture Notes in Computer Science* 6007:98–107.

Borgatti, S. P., M. G. Everett, and L. C. Freeman. 2002. *UCINET 6 for Windows: Software for social network analysis*. Cambridge, MA: Analytic Technologies.

Brewer, D. 1995. Cognitive indicators of knowledge in semantic domains. *Journal of Quantitative Anthropology* 5:107–28.

Garro, L. 2000. Remembering what one knows and the construction of the past: A comparison of cultural consensus and cultural schema theory. *Ethos* 28:275–319.

Hruschka, D. J., L. M. Sibley, N. Kalim, and J. K. Edmonds. 2008. When there is more than one answer key. *Field Methods* 20:315–37.

Johnson, J. C. 1990. *Selecting ethnographic informants*. Newbury Park, CA: Sage.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Romney, A. K. 1999. Cultural consensus as a statistical model. *Current Anthropology* 40:S93–S115.

Romney, A. K., S. C. Weller, and W. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88:313–38.

Ross, N. 2004. *Culture and cognition: Implications for theory and method*. Thousand Oaks, CA: Sage.

Weller, S. C. 1984. Consistency and consensus among informants: Disease concepts in a rural Mexican village. *American Anthropologist* 86:966–75.

_______. 1987. Shared knowledge, intracultural variation, and knowledge aggregation. *The American Behavioral Scientist* 31:178–93.

_______. 2007. Cultural Consensus Theory: Applications and Frequently Asked Question. *Field Methods*. 19:339–368.