Creating Shorter Scales to Measure Congregational Vitality.

By Linda Bobbitt, Congregational Vitality Project Researcher/Project Manager
Evangelical Lutheran Church in America
June 5, 2015

Abstract

After Bobbitt created the Congregational Vitality Survey (CVS), a need emerged for a shorter scale to be created and administered via annual parochial reports collected from every congregation by the denomination. Also, a shorter scale would have great utility when included with other instruments as a vitality subscale and used to create a self-scoring survey for congregational leadership. The original 29 item survey was effectively reduced to form a 15 item scale and a 5 item scale. The 15-item scale retained greater face validity, while the 5-item scale was the best predictor of external health. Both scales were internally consistent and explained a significant amount of variance in an external observation of a congregation's health.

Introduction

After Bobbitt[1] created the Congregational Vitality Survey (CVS), it began to be used by congregations primarily in four pilot synods. Synod and congregational leaders resonated with the three vitality scales: connecting with God, each other and the world. While these scales were helpful in expanding the concept of vital congregations beyond money and membership, the 29-item scale was too long to use across all settings. A need emerged for a shorter scale to be created and administered via annual parochial reports collected from every congregation by the denomination. Also, a shorter scale would have great utility when included with other instruments as a vitality subscale and used to create a self-scoring survey for congregational leadership.

Since the 29-item survey demonstrated high internal consistency and high inter-item correlations with recently tested samples, it suggested that some items duplicate each other conceptually. Therefore it is believed that a smaller scale could more efficiently provide the same or better reliability as a measure while still correlating with external observations of congregations.

The following criteria were establish to form the revised scale. It must:

- Represent each of the original scales conceptually. (Connections with God, each other and the world)
- Correlate well with the original scales aggregated (three scales combined into one average score) (This is to ensure that the shorter scale is a measuring the same thing as the longer scale.)
- Show equal or greater internal consistency (reliability) as the original scales.
- Demonstrate face validity for denominational, synodical and congregational leaders so that people will trust the shorter scale.
- Coincide with observed congregational health as determined by a multi-factor external rating system. Unless the scale scores connect to real life conditions, the scores are not helpful.

---

[1] Linda D Bobbitt, " Measuring Congregational Vitality: Phase 2 Development of an Outcome Measurement Tool," *Review of Religious Research* 56, no. 3 (2014).

Method

In 2013 and 2014, 37 ELCA congregations from six synods asked every member (who attended church within in a given week) to take the Congregational Vitality Survey (CVS). Frequently this survey was administered in conjunction with other services provided by the synod. This study used the individual responses to each item of the CVS averaged together to create congregational-level data. Congregational-level items were averaged to create scale scores.

Identifying Shorter Scales

In an effort to identify the most relevant items to include in a shorter scale, the external health rating (see methodology below) was correlated with each congregational level item. All but five of the 29 items were significantly correlated with the external rating and 20 of the 29 were correlated at $p < .01$ with correlations ranging between $r = .437$ (addressing social concerns) to $r = .664$ (sense of excitement about the future). Unfortunately, many of these items were also highly correlated with each other.

To narrow the field, items with significant correlations were selected that best represented face validity within each scale. Several groupings were created and tested with the following methodology:

1. Obtained correlations with the external health rating. Items with highest correlation were used first.
2. Assessed internal consistency of each group of items with Cronbach's alpha
3. Conducted factor analysis of each group using no rotation.
4. Performed linear regression to explore the relationship between small scales (independent variable) and the external health rating (dependent variable). Results show how much of the variation in the external health rating can be explained by each scale.

Priority was given to the groupings that best explained the external health ratings while still maintaining face validity.

Beyond a scale with strong face validity, I wanted to find a smallest scale possible that best predicted the external health rating. Different combinations of one to six items were tested using the same criteria and process but relaxing the expectation of face validity.

Multi-factor External Rating

Between three months and one year after congregations received synod services, researchers from the University of Northern Colorado's Social Research Lab called each congregation and interviewed both the pastor (clergy) and congregational president (lay person) if possible. A total of 63 interviews were done covering 44 congregations. Thirty-seven of those congregations also took the CVS. These were the congregation's used for the external rating.

All respondents were asked a variety of questions about the services they received from the synod and how they were using them now. In addition, they were asked to rate their congregational health and to identify their place on the congregational life cycle (Appendix A) both before and after the synod services were provided. Self-ratings were combined with external observations of the congregation including synod staff's ratings on health and life cycle scales, annually reported attendance and changes in attendance over the past two years, income/expenses, and the percentage of CVS respondents that

rated Overall Vitality as stable or better (as opposed to struggling). Unfortunately, not all information was available for every congregation.

All available information was considered by the author and used to create congregational health scores using the same A-D scale that was used in the interview. Most of the time the information was consistent across sources, however when ratings contradicted each other, the pastor's ratings were used as a base and modified by the other data.

Because the CVS was taken before the intervention in each instance, it was determined that the survey results should most closely match the pre-service scores for each congregation. Therefore, only the pre-service ratings were used to determine the external rating.

The following Health Scale was used by pastors, lay leaders, and synod staff as the congregational health rating. Note that these ratings include concepts associated with sustainability as well as missional vitality.

a. **Healthy and moving forward**
   The congregation has a clear sense of purpose and a commitment to mission. Overall health looks strong. The ministry is growing at a strong pace and is using models for ministry appropriate to continued effective outreach.

b. **Healthy but maintaining**
   The congregation appears basically healthy with no major health struggles, has some missional clarity but produces minimal results, is growing slowly or is now plateaued.

c. **Health concerns**
   The congregation exhibits some signs of health. Here there may be factors needing serious attention. Attendance is not increasing and resources are stretched. Without intentional change, the congregation is likely to struggle at its current level or shrink. Future use of resources should be evaluated carefully.

d. **Health in jeopardy**
   The congregation is clearly too small to maintain critical mass for mission or is in noticeable decline. The health of the ministry is also in question. A new strategy for ministry in the field is mandatory.

Even though the health categories are discrete, many respondents choose items between the letters (e.g. B+ or B-). To accommodate this distinction, numeric scores were assigned for each possible rating as shown in Table 1.

Table 1. Ratings for Health Scores

A+ = 4.25	B+ = 3.25	C+ = 2.25	D+ = 1.25
A = 4	B = 3	C = 2	D = 1
A -= 3.75	B -= 2.75	C- = 1.75	D- = .75

The assigned numeric scores were used as the dependent variable in the linear regression models.

To better understand how to interpret an individual congregation's scale score, a cluster analysis was performed using the Ward method. An ANOVA was performed to identify whether the clusters showed significant differences. (e.g., Is a congregation with high scale scores really different than one with middle or lower scale scores?). Upon significant differences, the Tukey post hoc test was performed to determine where significant differences existed and where they did not.

Results

Appendix B lists all item abbreviations and full text from the original CVS as well as which items are included in each of the new shorter scales. All items were rated on a 1 (lowest) to 5 (highest) scale and averaged to create a single composite score.

The initial short scale included four items from each of the original scales and a summary item. Leaders from four congregations and staff from three synods agreed that the items provided sufficient face validity. However, when the initial scale was then shared with denominational leaders, they believed that more items needed to be added for face validity (indicated in bold).

God: Future, Mission, **DeepGod**, Nurture, TryNew

Each Other: Incnew, Seekgift, Buildrel, MrgDis

World: PosForce, SocConc, LocalCom, Evang, **Justice**

Overall LivFaith

These scale items demonstrated high internal consistency r = .951**. The composite score ranged from 2.9 to 4.5, a mean of 3.9 and standard deviation of .397. Appendix C shows the inter-item correlations.

The results of the factor analysis showed that items from the Each Other scale were distinct from other items, particularly those on the world scale. This finding was consistent with scale performance in the original survey. Table 2 shows the factor loadings.

Table 2. Factor Analysis of 15 item
Parochial Report scale

	Component	
	1	2
Future	.839	.135
Mission	.909	-.240
TryNew	.793	.351
Incnew	.669	.422
Buildrel	.626	.729
MrgDis	.388	.848
SocConc	.778	-.436
Evang	.917	-.015
LocalCom	.780	-.478
Nurture	.967	-.079
PosForce	.822	-.401
Seekgift	.889	.258
Justice	.772	-.445
LivFaith	.954	.113
DeepGod	.840	-.121

Extraction Method: Principal
Component Analysis.

a. 2 components extracted.

Creating the Smallest Possible Scale (Mini CVS)

Twelve combinations including from one to six items were tested to find the one that best represented the external health rating. The result was a five-item scale.

God: Future, Trynew

Each Other: SeekGift

World: PosForce

Overall: LivFaith

This scale resulted in less face validity because it did not contain items about faith formation, missional attitudes, worship, evangelism or other areas important to vitality. The scale still worked because the items below correlated highly with the missing items (see Appendix C) without correlating as highly with each other. Therefore they covered more conceptual ground with less overlap.

These items demonstrated internal consistency at r = .920. Congregational scores ranged from 2.8 to 4.5, with a mean 3.9 and standard deviation of .44. Inter-item correlations are shown in Table 3.

Table 3. Mini-scale inter-item correlations

Inter-Item Correlation Matrix

	Future	PosForce	TryNew	Seekgift	LivFaith
Future	1.000	.661	.741	.725	.772
PosForce	.661	1.000	.562	.660	.711
TryNew	.741	.562	1.000	.751	.786
Seekgift	.725	.660	.751	1.000	.885
LivFaith	.772	.711	.786	.885	1.000

Table 4 shows factor analysis results. Here only one factor was found.

Table 4:.Mini-scale factor analysis

	Component
	1
Future	.882
PosForce	.806
TryNew	.871
Seekgift	.914
LivFaith	.944

Extraction Method: Principal Component Analysis.

Consistency with Previous Scale

Table 5. Inter-scale correlations

	Tiny 8	Parochial Report (averaged across)	All original items together
Tiny 8	1	.973**	.945**
Parochial Report		1	.967**
Quick Quiz			.969**
All Items			1

Table 5 shows high correlations among the three scales suggesting that both smaller scales are good representations of the original scale.

Comparing Scales with External Rating

Figure 1 shows a histogram with frequencies of congregational external health ratings. Congregations were distributed normally with most congregations were rated B followed by C.
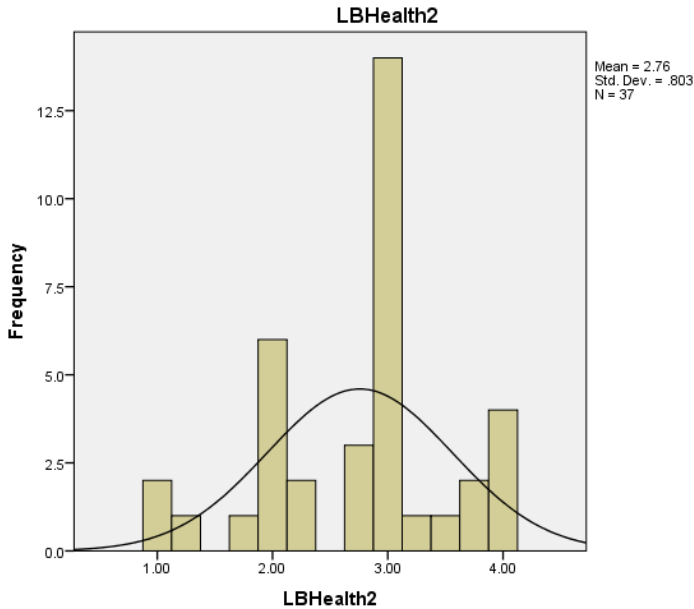
Figure 1. External Rating Histogram

Linear regression was used to explore the relationship between small scales (independent variable) and the external health rating (dependent variable). Table 6 shows the adjusted R^2 value, standardized and unstandardized Beta coefficients for the two smaller scales and the original scale items.

Table 6. Linear regressions between scales and external health score

	Mini-Vitality	Parochial Report (Averaged across)	All original items together
Adjusted R^2 & Standard Error (SE)	.501 SE .31332	.412 SE .30471	.371 SE .29547
Standardized Beta & t	.717** t=6.092**	.654** t=5.120**	.624** t=4.719**
Unstandardized Beta Coefficients & SE	.396 SE .065	.324 SE .063	.290 SE.061

**$p<.000$

While all scales explain a significant amount of the variance in the external health rating, the smallest scale does the best job by explaining 50% of the variance in health scores. The high adjusted R^2 value is surprising since the external health scores took factors like attendance and fiscal stability into account while the survey did not.

Scatter plots for each scale are displayed Figure 2.
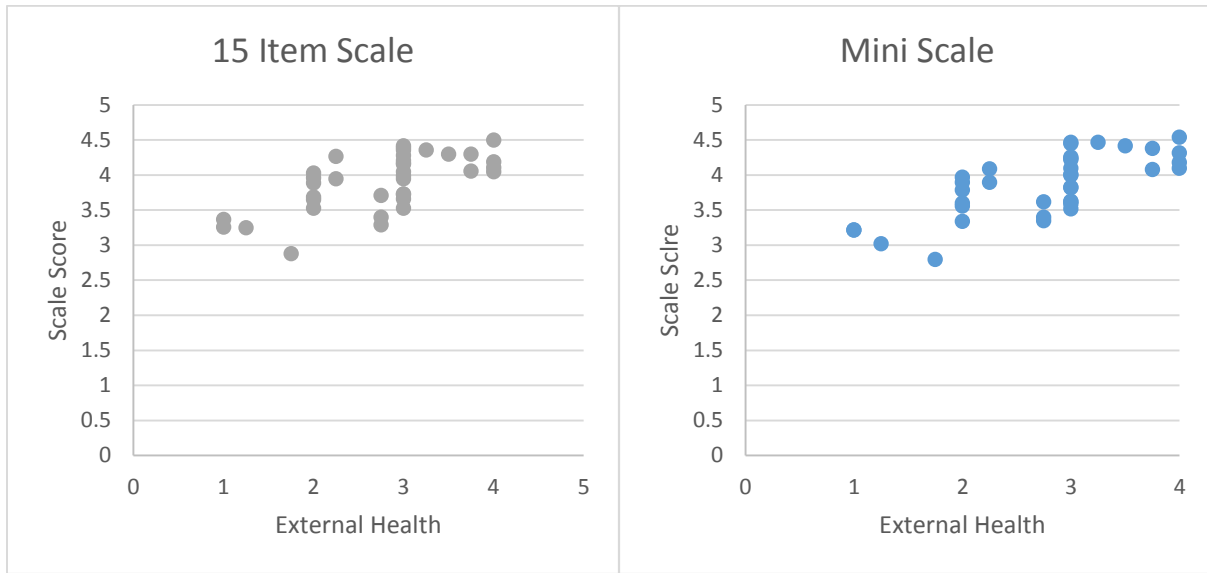
Figure 2. Scatter plots of subscales

As the scatter plots show, congregations were best distinguished between low and high, but that there is considerable overlap in mid-level scores.

Cluster analysis labeled each congregation in groups in four clusters according to the mini-scale scores (low to high). Table 7 shows the descriptive information for each cluster.

Table 7. Cluster descriptives

Cluster	N	Mean Mini-CVS score	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1	7	3.1936	.21456	.08110	2.9951	3.3920	2.80	3.40
2	12	3.6985	.13959	.04030	3.6098	3.7872	3.52	3.90
3	12	4.1309	.11326	.03269	4.0590	4.2029	3.97	4.32
4	6	4.4559	.05506	.02248	4.3982	4.5137	4.38	4.54
Total	37	3.8661	.44344	.07290	3.7182	4.0139	2.80	4.54

ANOVA was performed to examine the differences in external health ratings among these clusters and found significant differences (F = 11.00, p < .000.) Each cluster coincided with mean health scores which

represented the following approximate letter grades: Cluster 1= 1.79 (C-), Cluster 2 = 2.58 (C+), Cluster 3= 3.17 (B), and Cluster 4=3.42 (B+).

A Tukey post hoc test showed that the clusters were not all significantly different from each other. Cluster 1 was significantly lower than the other clusters. Clusters 2 and 3 were not significantly different from each other, and Clusters 3 and 4 were not significantly different from each other. Yet, Cluster 4 was significantly higher than Cluster 2.

Conclusion

The 29-item Congregational Vitality Survey was effectively reduced to form smaller scales which maintained face validity, internal consistency and construct validity. The 15-item scale retained greater face validity, while the 5-item scale was the best predictor of external health.

Both scales explained a significant amount of variance in an external observation of a congregation's health. This suggests that using these scales (particularly the mini scale) can yield a reasonable estimate of a congregation's overall vitality and sustainability. But caution is advised. While scale scores under 3.0 suggest that a congregation is probably struggling and scale scores over 4.4 suggest that a congregation is probably maintaining or thriving, there is considerable overlap in the scales, particularly when scale scores are between 3.4 and 4.4. Therefore, scale scores should be used as initial observations, baseline data collection for synods and denominations (with hundreds or thousands of congregations) and conversation starters. They should not be used for making high-stakes decisions about congregations.

All data used in this analysis was collected from an every-member style congregational survey providing a representative sample within each congregation. By using congregational means, sometimes wide variations among individual respondents were muted. Moving forward, the 15-item scale is scheduled to be used on annual parochial reports beginning in December 2015. It will also be provided to congregations as a self-scoring survey for congregational leadership beginning in the summer of 2015. Both of these applications will survey a small, unrepresentative samples of congregations (key informants). That will create a much broader range of scores which will need to be interpreted with care. Additional reliability and validation analysis will be needed after the scale is used nationally as a key informant survey.

Both of these subscales are made available for other denominations, middle judicatories and congregations to use. Feedback to the ELCA Research and Evaluation Department about the use and outcomes of this use is appreciated.

Bibliography

Bobbitt, Linda D. " Measuring Congregational Vitality: Phase 2 Development of an Outcome Measurement Tool." *Review of Religious Research* 56, no. 3 (2014): 467-84.

Congregational Cycle
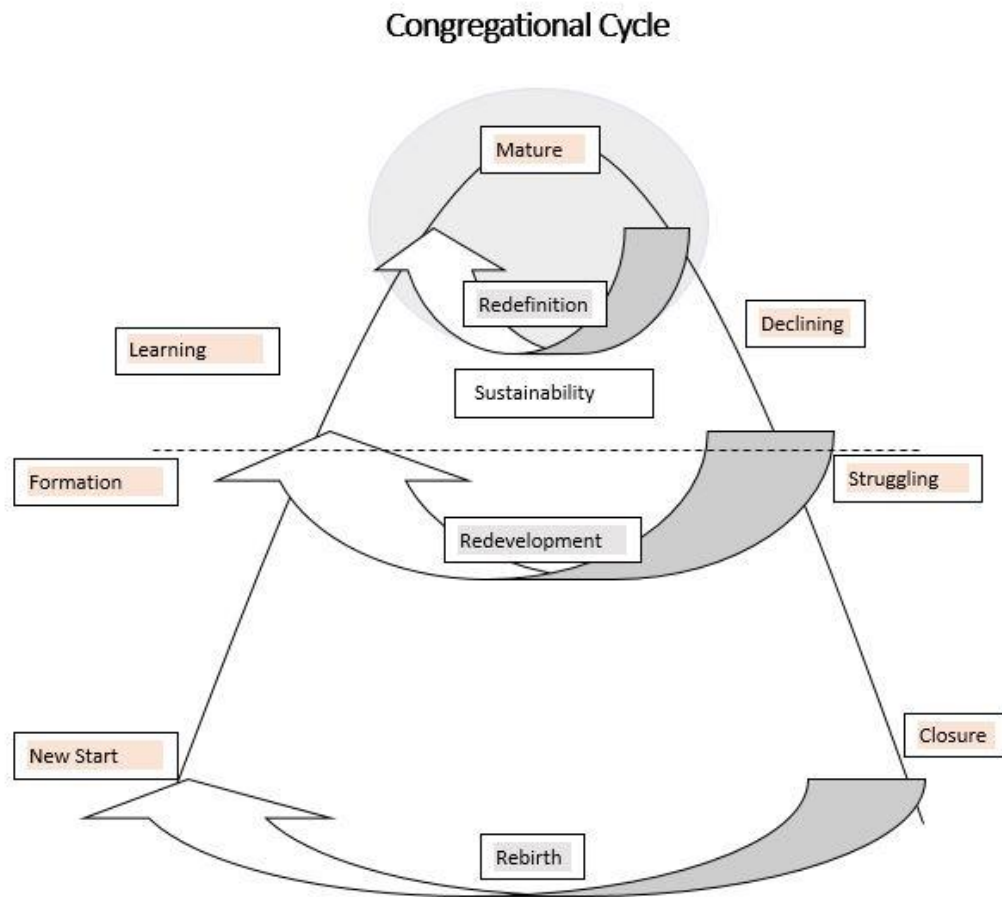
New Start (A)	Formation (B)	Learning (C)
New worshiping community	Organized but not yet self-sustaining	Self-sustaining, growing and developing programs and systems
Mature (D)	**Declining (E)**	**Struggling (F)**
Cycle through growth/change, stable, plateau, eval.,redefine, grow/change	Stuck in old patterns, loss of focus on mission, loss of members/money.	Same as declining but no longer sustainable (money or people or both).
Closure (G)		
Not enough resources or energy to continue or redevelop. Only option is to choose how to close.		

Appendix B. Survey Items

	Abbreviation	Full Survey	15 item	Mini
How well do these phrases describe your congregation? (*Responses: Poorly, Somewhat, Great*)				
This place is spiritually vital		X		
A clear sense mission	Mission	X	X	
Excitement about the congregation's future	Future	X	X	X
This place helps people deepen their relationship with God	DeepGod	X	X	
This cong. works for social justice/advocacy	Justice	X	X	
This cong. is a close-knit family		X		
A positive force in the community	PosForce	X	X	X
This cong. is focused on doing God's work		X		
This place has lots of meaningful activities		X		
Always ready to try something new	TryNew	X	X	X
How well do the following phrases describe the worship service you attend most often? (*Responses: Poorly, Somewhat, Great*)				
Welcoming to visitors		X		
Filled with a sense of God's presence		X		
Joyful		X		
Nurturing of people's faith	Nurture	X	X	
How well does your congregation… (*Responses: Hardly at all, Well, Very Well*)				
Accepting new members		X		
Incorporate new members into congregational life	Incnew	X	X	
Seek out and using the gifts of members of all ages	Seekgift	X	X	X
Build strong, healthy relationships among members	BuildRel	X	X	
Manage disagreements in a healthy respectful manner	MrgDis	X	X	
Involving youth/young adults in decision-making		X		
Helping children grow in their faith		X		
Helping adults grow in their faith		X		
Praying (during and outside worship services)		X		
Address social concerns (helping those in need)	SocConc	X	X	
Equip members to share their faith with others	Equip	X	X	
Interact with the local community		X	X	
Partnering with other cong. in area		X		
Sharing information about the ELCA or synod		X		
Help members live out their faith in daily lives	LivFaith	X	X	X

Appendix C: Parochial Report Intra item correlations

		Future	Mission	DeepGod	Justice	Pos Force	TryNew	Nurture	Incnew	Seekgift	Buildrel	MrgDis	Soc Conc	Evang	Local Com	Liv Faith
Future	Pearson Correlation	1	.773**	.738**	.526**	.661**	.741**	.820**	.489**	.725**	.620**	.433**	.470**	.683**	.631**	.772**
	Sig. (2-tailed)		.000	.000	.001	.000	.000	.000	.002	.000	.000	.007	.003	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Mission	Pearson Correlation	.773**	1	.778**	.765**	.789**	.641**	.881**	.440**	.729**	.379*	.161	.785**	.857**	.834**	.865**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.006	.000	.021	.342	.000	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
DeepGod	Pearson Correlation	.738**	.778**	1	.771**	.684**	.534**	.853**	.579**	.679**	.471**	.188	.680**	.651**	.621**	.741**
	Sig. (2-tailed)	.000	.000		.000	.000	.001	.000	.000	.000	.003	.265	.000	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Justice	Pearson Correlation	.526**	.765**	.771**	1	.755**	.371*	.786**	.478**	.582**	.154	-.050	.893**	.645**	.679**	.645**
	Sig. (2-tailed)	.001	.000	.000		.000	.024	.000	.003	.000	.363	.770	.000	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
PosForce	Pearson Correlation	.661**	.789**	.684**	.755**	1	.562**	.800**	.338*	.660**	.225	.001	.764**	.729**	.903**	.711**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.041	.000	.181	.997	.000	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
TryNew	Pearson Correlation	.741**	.641**	.534**	.371*	.562**	1	.688**	.533**	.751**	.709**	.631**	.456**	.749**	.540**	.786**
	Sig. (2-tailed)	.000	.000	.001	.024	.000		.000	.001	.000	.000	.000	.005	.000	.001	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Nurture	Pearson Correlation	.820**	.881**	.853**	.786**	.800**	.688**	1	.646**	.851**	.555**	.269	.772**	.879**	.761**	.913**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.107	.000	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Incnew	Pearson Correlation	.489**	.440**	.579**	.478**	.338*	.533**	.646**	1	.723**	.704**	.552**	.410*	.563**	.199	.646**
	Sig. (2-tailed)	.002	.006	.000	.003	.041	.001	.000		.000	.000	.000	.012	.000	.237	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37

Seekgift	Pearson Correlation	.725**	.729**	.679**	.582**	.660**	.751**	.851**	.723**	1	.753**	.506**	.518**	.808**	.565**	.885**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000		.000	.001	.001	.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Buildrel	Pearson Correlation	.620**	.379*	.471**	.154	.225	.709**	.555**	.704**	.753**	1	.830**	.192	.549**	.133	.664**
	Sig. (2-tailed)	.000	.021	.003	.363	.181	.000	.000	.000	.000		.000	.256	.000	.433	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
MrgDis	Pearson Correlation	.433**	.161	.188	-.050	.001	.631**	.269	.552**	.506**	.830**	1	-.006	.342*	-.080	.466**
	Sig. (2-tailed)	.007	.342	.265	.770	.997	.000	.107	.000	.001	.000		.973	.038	.637	.004
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
SocConc	Pearson Correlation	.470**	.785**	.680**	.893**	.764**	.456**	.772**	.410*	.518**	.192	-.006	1	.730**	.735**	.686**
	Sig. (2-tailed)	.003	.000	.000	.000	.000	.005	.000	.012	.001	.256	.973		.000	.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
Evang	Pearson Correlation	.683**	.857**	.651**	.645**	.729**	.749**	.879**	.563**	.808**	.549**	.342*	.730**	1	.770**	.943**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.038	.000		.000	.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
LocalCom	Pearson Correlation	.631**	.834**	.621**	.679**	.903**	.540**	.761**	.199	.565**	.133	-.080	.735**	.770**	1	.702**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.001	.000	.237	.000	.433	.637	.000	.000		.000
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37
LivFaith	Pearson Correlation	.772**	.865**	.741**	.645**	.711**	.786**	.913**	.646**	.885**	.664**	.466**	.686**	.943**	.702**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.004	.000	.000	.000	
	N	37	37	37	37	37	37	37	37	37	37	37	37	37	37	37

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).