http://dx.doi.org/10.7468/jksmed.2014.18.1.75
J. Korean Soc. Math. Educ., Ser. D, Res. Math. Educ.
Vol. 18, No. 1, March 2014, 75–87

How to Develop a Scale Measuring an Affective Construct in Mathematics Education Research

RYANG,Dohyoung
Department of Mathematicsand Statistics, Univ. of North Carolina at Greensboro, Greensboro, NC 24702, USA; Email: dryang@uncg.edu

(Received January 7, 2014; Revised March 17, 2014; Accepted March 21, 2014)

It is central to use a scale to measure a person's level of a construct in mathematics education research. This article explains a practical process through which a researcher rapidly can develop an instrument to measure the construct. The process includes research questioning, reviewing the literature, framing a background theory, treating the data, and reviewing the instrument. The statistical treatment of data includes normality analysis, item-total correlation analysis, reliability analysis, and factor analysis. A virtual example is given for better understanding of the process.

*Keywords:*efficacy, factor analysis, mathematics, preservice teacher, validity
MESC Classification: B59
*MSC2010Classification:*97C70

INTRODUCTION

Many research studies in mathematics education may require assessing a person's level of a psychological construct attributed in the affective domain such as emotion, beliefs, value and attitudes. Many research studies evidenced that an affective construct of a person influences on his or her actual ability to accomplish the related task. A multiple-choice testassessing a person's knowledge (or intelligent ability) has only two outcomes: correct or incorrect answers. Different from a knowledge test, measurement of a psychological affective construct must use a rating scale, for example, 5-point rating scale with options: Strongly Disagree, Disagree, Uncertain, Agree, and Disagree. Recently, there has been controversy that five options are not enough to reflect the whole continuum in human minds (Kislenko&Grevholm, 2008). Some researchers use 7 or even 9 point rating scale; other researchers still count a 5-point rating scale as good to use. This article explains a general processof developing an instrument measuring an affective constructin mathematical education research. The process of the development includes problem posing, literature review, theoretical framework, data collection, instrumentation, data analysis, results and review of the instrument. A virtual example is provided for better sense to the readers, where the construct is mathematics teaching efficacy beliefs and a 5-point scale is used just for convenience.

POSING A PROBLEM

Theresearch question is to develop a scale measuring a construct. A researcher should pose the construct that he or she like to measure. Since a construct could be influenced bythe level of the population and/or the settings surrounding the population, the researcher should specify the level of the population, for instance high school students or preservice teachers. In the virtual example, the construct is mathematics teaching efficacy beliefs and the researcher is interested in preservice teachers rather than in-service teachers in Korea so the research question is to develop a scale measuring mathematics teaching efficacy of Korean preservice teachers. Also, the researcher should explain why the study is important in the context of given settings. Thus, the literature review is critical in that theresearchercan justify the value of the study he or she is conducting.

REVIEWING LITERATURE

Throughout the review of literature on the construct, theory, and related research studies, a developer can present the definition of the construct, importance of the construct, previous studies, some issues on the construct, and so forth. Then must tell the originality of the solution to the posed research question. Literature review in brief supporting the virtual example follows in the next paragraph.

Teacher efficacy refers to a teacher’s self-perceived beliefs regarding his/her ability to organize and execute courses of actions to successfully accomplish a specific teaching task in a particular context (Tschannen-Moran& Hoy,2001). Since teacher efficacy beliefs are context specific and subject-matter specific (Tschannen-Moran,Woolfolk Hoy& Hoy, 1998), mathematics teaching efficacy can powerfully predict the teachers’ future behavioral actions in teaching mathematics as well as influence student mathematics motivation and outcome(Barr, 2005; Mojavezi&Tamiz, 2012; Swars, 2005; Swars, Smith, Smith, & Hart, 2009; Utley, Bryant & Moseley, 2005), and negatively related to mathematics anxiety (Gresham, 2008). In those studies, Enoch, Smith &Huinker’s (2000) Mathematics Teaching Efficacy Beliefs Instrument (MTEBI) was used.

Despite the fame of the MTEBI in mathematics education research, the factorial validity may be questionable. Gibson and Dembo's (1984) Teacher Efficacy Scale is a prototype measure to teacher efficacy instruments. The factor structure of Gibson and Dembo's scalewas controversial among researchers. Brouwers and Tomic (2003) tested various factor models of 2-, 3-, and 4-factor models; they found that all aforemenstionedfactor structures did not reach a sufficient level of factorial validity. Theresult indicated that Gibson and Dembo's scale was not suitable for a research study anymore. The MTEBI, as a variation of Gibson and Dembo's scale, is potentially to have weakness in factorial validity.

Though the MTEBI is considered valid for measuringmathematics teaching efficacy in the United States, it is no guarantee that the MTEBI works well in other cultures. Since teacher efficacy may vary culture to culture (Lin, Gorrell, & Tayler, 2002), it is suggested that an experimental study should validate the use of the instrument in that culture. There are controversial results in the studies. Alkhateeb (2004) validated the MTEBI for Arabic speaking Jordan preservice teachers while Cakiroglu(2008) and Chang (2003)'s pilot test reported that the second factor has low reliability. Thus, there still exists blurriness about the factorial validity of the MTEBI. This study is original in that the blurriness will be cleared in the newly developed measure, the Virtual Mathematics Teaching Efficacy Scale (VMES), in the virtual example.

FRAMEWORKING

Also, a construct must be rooted in a psychological theory. Usually, there existmultiple theories within which the construct can be explained. The developer will find such a theory throughout the literature. This process is called theoretical (or conceptual) framework. Review of literature should include this process. For the virtural example, a brief introducation to a theory that can explain teachers' efficacy beliefs is given here in the next paragraph.

Bandura (1997) conceptualized self-efficacy as a media tri-actively working among stimuli, responses and a person's cognitive process, in his social cognitive theory of learning.Bandura argued that efficacy beliefs are well explained in two dimensional model of self-efficacy (or personal efficacy, PER) and outcome expectancy (OUT). The construct in the virtual example is mathematics teaching efficacy. According to Bandura's theory, the variables explaining this construct are the PER and the OUT in mathematics teaching. These should serve as the two subscales of the instrument. The PER deals with personal sense of effectiveness in mathematics teaching while the OUT describes general beliefs about the effect of the social/cultural settings surrounding mathematics teaching.

COLLECTING DATA

The sample should be select randomly in an experimental study. Randomness is simple but hard to accomplish. The virtual example is presented here with assumption of random sampling. Data were collected from 12 national universities of education in South Korea. Program directors (or department chairs) gave informed consent for conducting the study. The survey was distributed to the participants in a regular class, and was returned in 20 minutes. The survey packet consists of the three sections: the cover letter (including directions and consent information), demographic questionnaire (e.g., age, gender, and class level), and the mathematics teaching efficacy instrument. The total participants,after deleting 15 incomplete cases, were 1000 elementary preservice teachers. Among them, 600 (60%) were female and 400 (40%) were male; 200 (20%) freshmen, 300 (30%) sophomores, 300 (30%) juniors, and 200 (20%) seniors. Average age was 22.5 (SD = 3.00) years old.

INSTURMENTATING

The development of a scale is nothing but the process of deleting weak items so a modified model can fit well the theoretically assumed model. By doing so, the scale's reliability and validity will increase. The length of an instrument is an issue because if an instrument is too long, the examinee cannot keep concentrating on the survey so the scale's reliability is harmfully affected. Researchers (cf. Bae, 2006) suggested that about 4 to 8 items in a scale are appropriate, and an instrument would have 20 items or so within a few subscales. Then, a survey packet of such instrument will be returned in only 10 minutes. In this way, the outer factors having effect on the scale reliability would be critically increased. In the virtual example of this article, the VMES will have two scales, the PER and the OUT. Thus, about 16 to 18 items are desirable in the final model. The first form would have items three times the final form. In the virtual example, the first form has 54 items (32 PER items and 22 OUT items).

Among 54 items, the first 21 items were the revised MTEBI items for Korean preservice teachers (Ryang, Thompson, Shwery, 2011). The PER items are stated in the first person and the future tense, for instance, "I will be able to answer students' mathematics question in class." While the OUT items are stated in the third person and the present tense, for instance, "When a student does better than usual in mathematics, it is because the teacher exertedextra effort." The next 10 items were suggested by Korean mathematics teacher educators. For example, "I will teach mathematics in such a way that they advance in a mathematics assessment.”The next 10 items were obtained from the modification (exchanging the PER and OUT wording) of the MTEBI items, e.g., “I will give extra effort to an underachieved student” which is an OUT item as shown previously. The next 7 items are from the other instruments or from the literature, for instance, “I will implement well a new math teaching strategies in class.” All OUT items odd numbered till 43 while all PER items were even numbered till 44 and 45 to 54.

PER: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54
OUT: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44

ANALIZING DATA

For developing the VMES, first, the item was coded by a letter P or O (depending on the variables) followed by the item number, i.e., there are two types of items, P-items and O-items. Then, the response of an item was coded by 1 for Strongly Disagree to 5 for Strongly Agree. On the whole data set (N = 1000), the normality of an item variable would be tested. After that, the data set was divided into the two subsets each of which different types of factor analysis would be conducted. The factor structure is *explored* in the first data set (N = 419), and the structure is then *confirmed* in the second data set (N = 500). Testing the factor structure by way of different methods in different data sets will increase cross-validity. This study used the SPSS 21 program for the normality test and the exploratory factor analysis and the LISREL 8.80 program for the confirmatory factor analysis.

WRITING RESULTS

The results in this section are not written from the actual data but stated virtually in order to make more sense to the reader on the steps in developing the scale (or in deleting items in a reasonable sense).

Normality Test

A random variable’s normal distribution can be determined by the degree of skewness and kurtosis which are sort of horizontal and vertical violation from a normal distribution. The LISREL program provided the translated Z-scores of the combination of the skewness and kurtosis for an item variable and its p-value. In thevirtual example, Table 1 showed some descriptive statistics like mean, standard deviation, skewness-kurtosis *Z*-score, and *p*-value for the first six items.An item with the*p*-value less than .05 are considered as violating the normality. In the virtual example, 25 items in the PER and 10 items in the OUT, with together 35 items passes the normality test.

Table 1. Mean, standard deviation, and significance for normality (in part)

Item	Mean	S. D.	*Z*-score	*p*-value
P1	3.721	0.798	2.325	0.313
O2	3.976	0.782	5.355	0.069
P3	3.399	0.856	0.049	0.976
O4	3.454	0.853	4.599	0.125
P5	2.878	0.898	0.751	0.687
O6	3.529	0.884	0.943	0.518

Exploratory Factor Analysis

The Principal Component Analysis (PCA) with promax rotation explored the 2-factor structure in the 35 items that passed the normality test. The 25 P-items and the 10 O-items are respectively assumed to have a 1-factor structure. Then, the combined scale would have a 2-factor structure.

First, a 1-factor structure of the PER scale was explored. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .93; Bartlett's test of sphericity was significant (χ^2= 3500.123, df = 300, p < 0.001). The PCA with promax rotation on the 25 P-items initially extracted 5 components with eigenvalues greater than 1. But, the first component had a distinctively high eigenvalue and others made a flat land smoothly decreasing to the right in the scree plot. A 1-factor structure was assumed on the PER scale.

Next, a 1-factor structure of the OUT scale was then explored. KMO measure was .765. Bartlett's test of sphericity was significant (χ^2=650.567, df = 45, p < 0.001). The PCA with promax rotation on the 10 O-items initially extracted 2 components with eigenvalues greater than 1. The scree plot suggested that the scale might have a 2-factor structure rather than a 1-factor structure. The 2-factor solution PCA on the 10 O-items indicated that the three items constituted the minor factor. Deleting these items, the remaining 7 O-items were suggested to have a 1-factor structure.

Then, a 2-factor structure was explored on the combined scale ofthe 25 P-items and the 7 O-items. The KMO index was .915; Bartlett's sphericity test was significant (χ^2= 4440.440, df = 496, p < 0.001). The PCA with promax rotation on the 32 item model extracted six components with eigenvalues greater than 1. But, first two components had distinctively higher eigenvalues than the others of which eigenvalues made a smooth decrease appearing a flat tail to the right in the scree plot (see Fig. 1). A 2-factor structure was suggested on the combined scale (32-item model).

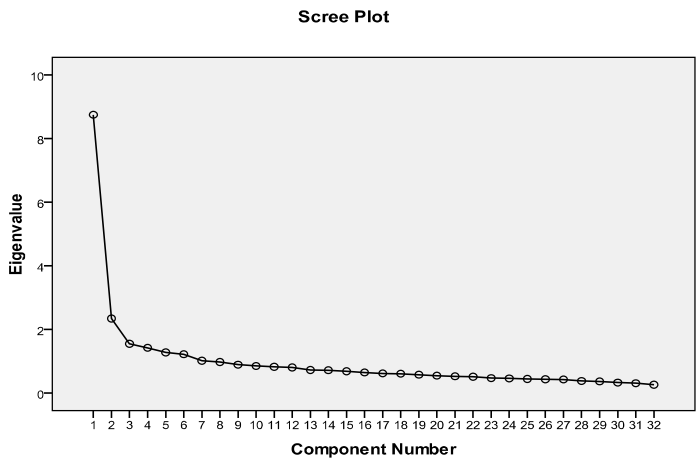

Fig. 1. The scree plot of the 32 items. A 2-factor structure is observed.

Thus, in order to look into the items' behaviors acting on the factors, the PCA with promax rotation was conducted with 2-factor solution on the 32-item model.All O-items were loaded to the component 2. However, all of P-items were not loaded to the component 1. TwoP-itemswere loaded to the component 2; seven items were double-loaded.These nine items hurted the scale's factorial validity very much.After deleting these nineP-items, the 23 items (16 P-items and 7 O-items) showed *the* 2-factor structure (see Table 2). The components 1 and 2 explained 23.23% and 11.77% of the total variance, respectively.From the factor loading, the two components 1 and 2 are the PER and the OUT, respectively.

Reliability Analysis

A weak item possibly exists on a scale to reduce the reliability of the scale to which the item belongs. These items will be removed sothe scale will have stronger reliability. To find such an item, internal consistency reliability(Cronbach's alpha) of each subscale in the 23-item modelwas compared to the alpha after an item was deleted (See Table 2). The PER scale (16 items) had $\alpha = .870$, the MTOE (7 items)$\alpha = .700$. The alpha after deleting P54 was greater than that before deleting the item. On the other way, an item's correlation to the whole scale serves as an index by which a developer can determine if the item would contribute to the whole scale measuring the construct. An item with lower ITC is considered as being deleted from the scale. The cut-off value depends on the developer and the settings for developing a scale. For the VMES in the virtual example, .32 was used. P54 has ITC = .312 < .32 indicating that this items does not contriute to the whole instrument to measure the construct. In addition, P54 has low factor loading (.330) though it is acceptable. Therefore, this item will be removed and the 22 items thus remain in the scale.

Table 2.PCA and Reliability on the 23 Items

Item	Component		ITC[a]	Reliability[b]
	1	2	Pearson *r*	*α*-if-item-deleted
P1	.700		.350	.864
P3	.695		.350	.866
P5	.690		.400	.862
P7	.685		.400	.860
P9	.680		.450	.853
P11	.650		.450	.869
P13	.500		.350	.859
P15	.450		.350	.854
P43	.520		.505	.866
P44	.535		.560	.861
P45	.565		.430	.863
P47	.490		.735	.861
P49	.480		.625	.864
P51	.417		.550	.865
P53	.389		.490	.858
P54	.330		.312*	.872**
O2		.565	.425	.698
O16		.630	.510	.691
O22		.525	.425	.696
O28		.610	.380	.698
O32		.410	.375	.686
O36		.526	.350	.690
O40		.545	.415	.693

[a]ITC = Item-Total Correlation; [b]PER scale reliability α = .870; OUT scale reliability α = .700
*The value less than .32 is too low to contribute to the correlation.
**The alpha if P54 deleted is greater than the alpha before the item.
Note. Factor loading values less than .32 were erased in the table.

Confirmatory Factor Analysis

Unidimensionality is a required assumption to conduct a confirmatory factor analysis. Actually, this assumption was tested on the process of exploratory factor analysis. The Structural Equation Modeling (SEM) provides a way to confirm an empirical model to fit the theoretical model. The SEM reports various fit indices for the model such as χ^2 with degree of freedom (df), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and goodness-of-fit index (GFI). When deleting a weak item one at a time, these indices would beimproved. The target values of these fit indices (Abell, Springer &Kamata, 2009; Bae, 2006; Hu &Bentler, 1999) were shown in the first row of Table 3.

Table 3. Comparison of fit indices among some the models

Model	χ^2	df	RMSEA	SRMR	CFI	GFI
Target Value	χ^2 / df<	3	.08	.05	.90	.90
22 Item Model	705.06	220	.070	.060	.80	.88
18 Item Model (A)	343.66	134	.055	.049	.88	.94
18 Item Model (B)	268.55	131	.045	.044	.90	.95

The SEM on the 22-item model over the second data set (N = 500) reported the fit indices shown in Table 3. Also, the SEM suggested modifying the model by adding a path from a P-item to the OUTvariable (e.g., P13 → OUT) or from an O-item to the PERvariable (e.g., O16 → PER) or between the two items with different factors (e.g., P15 ↔ O16). The SEM detected such weak items by reporting χ^2-decrease if a path is allowed between them. An item with the most χ^2-decreasewas deleted one at a time. After deleting four items (P13, P15, P44, O16),the 18 items (Model A) showed good index values but the CFI.Moreover, allowing covariance error existing within some P-items (P1 ↔ P3; P45 ↔ P47) will improve fitting to the theoretical model. Little covariance within the same variable happens commonly, which supportsthe fact that the two variables PER and OUT are intertwined each other in Bandura's theory.Thus, the final 18-item model (B) had better fitindices all of which met the targeted values.

The SEM also provides the structural coefficients between an item and the latent variable (factor loading) and between the latent variables. Different from exploratory factor analysis, the SEM provides the items' factor loadings with the errors and so the *t*-values.An item whose*t*-value less than 1.96 in a two-tailed test, with 0.05 significance level,does not help in constructing the scale's validity.In this sense, the SEM provides more rigorous way of factor analysis. All 18 items in the model had factor loadings with t>1.96 indicating that the P-items and O-items converge well to the PER and the OUT variables, respectively. On the other hand, the PER and the OUT should be discriminated because they are different variables. A way to investigate such discrimination is to test the null hypothesis that the two variables covariate completely, i.e., $\varphi = 1$ (Bae, 2006). In theVMES, the structural coefficient between the PER and the OUT was .31 with standard error 0.06. The φ-confidence interval with significance level .01 was 0.31 ± 2.58 (0.06) = (0.155, 0.465), which did not include 1. So, the null hypothesis was rejected and thus it cannot be said that $\varphi = 1$. Therefore, the discrimination between the PER and the OUT are secured in a reasonable degree.

Table 4.LISREL estimates (maximum likelihood) of the 18-item model

	MTPE			MTOE		
	Coefficient	Error	*t*-value	Coefficient	Error	*t*-value
P1	.34	.04	8.50			
P3	.41	.04	10.25			
P5	.42	.04	10.50			
P7	.28	.04	7.00			
P9	.39	.04	9.75			
P11	.40	.04	10.00			
P43	.37	.04	9.25			
P45	.41	.04	10.25			
P47	.58	.04	14.50			
P49	.43	.04	10.75			
P51	.36	.04	9.00			
P53	.48	.04	12.00			
O2				.30	.05	6.00
O22				.32	.05	6.40
O28				.37	.05	7.40
O32				.36	.05	7.20
O36				.25	.05	5.00
O40				.22	.05	4.40

REVIEWING THE DEVELOPMENT

In order to develop a new instrument to measure Korean preservice teachers' efficacy beliefs in teaching mathematics, the first form instrument of the 54 items was developed from an already existed instrument MTEBI, interview of Korean mathematics teacher educators, and literature review. All items were believed to ask a teacher's personal effectiveness in mathematics teaching and students' mathematics outcome expectation. Several statistical methods were used to test the scale's reliability and validity. During the process, weak items like violating the normality, reducing the scale's reliability, and unfitting the 2-factor structure were deleted. The final 18-item model, named the VMES, is believed as a good instrument to measure the construct in Korean contexts. Note that the VMES was developed in different statistical methods on different data sets. In this way, the instrument's cross-validity would be increased. The VMES consists of the two subscales: the PER (12 items) and the OUT (6 items). The reliability of these two subscales were .850 for the PER and .710 for the OUT.

CONCLUSION

There are lots of constructs in the affective domain such as emotion, value, beliefs, and attitudes that are important variables to students' successin learning mathematics and teachers'success in teaching mathematics. Also, each construct should be separately dealt according to the position of the population, e.g., teachers or learners. For example, a researcher may be interested in developing a measure for teachers' (or students') epistemological beliefs on mathematics. The level of maturity of subjects should also be considered, for instance, expert (or preservice) teachers' epistemology on mathematics. In addition, a construct in mathematics education is influenced by social and cultural shifts and educational settings surrounding the population in the time that the research study is conducted. So, the matter of reliability and validity is not a one-time completion job but ongoing process over and over a time period. Also, theoretical conceptualization and/or statistical strategies and skills are perhaps advanced in the future. Further studies should continue on establishing validity and reliability on an instrument. So, a research study will obtain fresh information even for the same construct.

This article describes, in contemporary sense, a way how to develop rapidly an instrument to measure a construct in mathematics education research. The development process, in general, includesproblem posing, justifying the research needs, literature review, seeking theory framing the study, developing the first form, collecting, coding, and analyzing data, writing results, and reviewing the process to the final model.Concrete explanation was given in a virtual example where the construct is mathematics teaching efficacy and the subjects were Korean preservice elementary teachers. This article, hopefully, will help a researcher to develop a measure for a psychological construct in mathematical education research studies.

REFERENCES

Abell, N.; Springer, D. W. & Kamata, A. (2009).*Developing and validating rapid assessment instruments*. Oxford University Press.

Alkhateeb, H. M. (2004). Internal consistency reliability and validity of the Arabic translation of the mathematics teaching efficacy beliefs instrument.*Psychological Reports* **94**.833–838.

Bae, B. (2006). *LISREL structural equation model: Understanding, practice, and programming*. Seoul: Chongram.

Bandura, A. (1997). *Self-efficacy*: The exercise of control. New York: W. H. Freeman.

Barr, L. C. (2005). *The effects of parent and teacher self-efficacy on the academic achievement motivation of preschool children.Unpublished master's thesis*.Winston-Salem, NC:Wake Forest University.

Brouwers, A. & Tomic, W. (2003).A test of the factorial validity of the Teacher Efficacy Scale.*Research in Education* **69**, 67–79.

Cakiroglu, E (2008). The teaching efficacy beliefs of pre-service teachers in the USA and Turkey, *Journal of Education for Teaching***34(1)**, 33–44.

Chang, Y. L. (2003). *An examination of knowledge assessment and self-efficacy ratings in teacher preparation programs in Taiwan and the United States*. Unpublished Doctoral Dissertation, Moscow, ID:University of Idaho.

Enochs, L.; Smith, P. & Huinker, D. (2000).Establishing factorial validity of the mathematics teaching efficacy beliefs instrument.*Sch. Sci. Math*.**100(4)**, 194–202. ME **2006c.**01495

Gibson, S. & Dembo, M. H. (1984). Teacher efficacy: A construct validation*. Journal of Educational Psychology* **76**, 569–582.

Gresham, G. (2008). Mathematics anxiety and mathematics teacher efficacy in elementary preservice teachers.*Teaching Education* **19(3)**, 171–184.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* **6(1)**, 1–55.

Kislenko, K. &Grevholm, B. (2008). The Likert scale used in research on affect: A short discussion of terminology and appropriate analyzing methods. Paper presented at the 11th International Congress on Mathematical Education at Monterey, Mexico on July 6–13, 2011. http://tsg.icme11.org/document/get/415

Lin, H., Gorrell, J. & Taylor, J. (2002). Influence of culture and education on U. S. and Taiwan preservice teachers' efficacy beliefs. *Journal of Educational Research* **96(1)**, 37–46.

Mojavezi, A. &Tamiz, M. P. (2012).The Impact of Teacher Self-efficacy on the Students' Motivation and Achievement.*Theory and Practice in Language Studies* **2(3)**, 483-491.

Ryang, D., Thompson, T. &Shwery, C. (2011). Analysis of Korean mathematics teacher educators' response to the Mathematics Teaching Efficacy Beliefs Instrument. *J. Korea Soc. Math. Educ. Ser. D***15(3)**, 229–250.

Swars, S. L. (2005). Examining perceptions of mathematics teaching effectiveness among elementary preservice teachers with differing levels of mathematics teacher efficacy.*Journal of Instructional Psychology***32(2)**, 139–147.

Swars, S. L.; Smith, S. Z.; Smith, M. E. & Hart, L. C. (2009). A longitudinal study of effects of a developmental teacher preparation program on elementary prospective teachers' mathematics beliefs.*J. Math. Teach. Educ*.**12(1)**, 47–66. ME **2009f.**00153

Tschannen-Moran, M. & Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education* **17**, 783–805.

Tschannen-Moran, M.; Woolfolk Hoy, A. & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research* **68(2)**, 202–248. http://mxtsch.people.wm.edu/Scholarship/RER_TeacherEfficacy.pdf

Utley, J.; Bryant, R. & Moseley, C. (2005).Relationship between science and mathematics teaching efficacy of preservice elementary teachers.*Sch. Sci. Math.***105(2)**, 82–87. ME **2007e.**00070