Practical Assessment, Research, and Evaluation

2016

Partial Least Squares Structural Equation Modeling with R

Hamdollah Ravand

Purya Baghaei

Partial Least Squares Structural Equation Modeling with R

Hamdollah Ravand, *Vali-e-Asr University of Rafsanjan and University of Jiroft, Iran*
Purya Baghaei, *English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran*

Structural equation modeling (SEM) has become widespread in educational and psychological research. Its flexibility in addressing complex theoretical models and the proper treatment of measurement error has made it the model of choice for many researchers in the social sciences. Nevertheless, the model imposes some daunting assumptions and restrictions (e.g. normality and relatively large sample sizes) that could discourage practitioners from applying the model. Partial least squares SEM (PLS-SEM) is a nonparametric technique which makes no distributional assumptions and can be estimated with small sample sizes. In this paper a general introduction to PLS-SEM is given and is compared with conventional SEM. Next, step by step procedures, along with R functions, are presented to estimate the model. A data set is analyzed and the outputs are interpreted.

Structural equation modeling (SEM) is a technique which combines factor analysis and regression. As compared to conventional statistical techniques such as regression, SEM is a more robust approach to testing substantive theories. Psychological constructs are usually complex and many variables might combine and interact to affect any phenomenon simultaneously. However, statistical techniques such as regression make simplifying assumptions about complex phenomena by considering only a limited number of variables to explain the variance in a dependent variable. These techniques do not further assume any relationships among the independent variables. According to Schumacker and Lomax (2004), SEM is more compatible with what happens in real-life because (a) it takes into account the relationships among many variables simultaneously and (b) in contrast to techniques such as regression which assume the measurement of the variables is error-free, SEM takes measurement error into account. SEM can simultaneously examine relationships among observed variables and latent variables as well as among latent variables.

Appealing as SEM may be, some of its assumptions are hard to meet in some research contexts. In terms of data distribution, it requires normally distributed data. As to sample size, relatively large sample sizes are required. Alternatively, Stevens (1992) suggests a sample size of fifteen participants per observed variable. Lohilen (1992) argued that researchers should choose at least one hundred participants for measurement models with two to four factors.

There are two types of SEMs: the conventional SEM which is referred to as *covariance-based* SEM (CB-SEM) and the partial least squares SEM (PLS-SEM) which is *variance based*. Due to the estimation procedures employed in each of the two types of SEM, they make different distributional assumptions and aim at different objectives. Using maximum likelihood (ML) estimation method, CB-SEM aims at estimating model parameters so that the discrepancy between the model-implied and sample covariance matrices is minimized. ML estimation implies a multivariate normal distribution. On the other hand, PLS-SEM, originally developed by Wold (1966, 1982, 1985) and Lohmoeller (1989), aims at maximizing dependent variables' explained variance by adopting an ordinary least squares estimation method. PLS-SEM is a nonparametric method hence makes no distributional

Practical Assessment, Research, and Evaluation, Vol. 21 [2016], Art. 11

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 2
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

assumption. CB-SEM is more suited to well-researched domains where enough theoretical and substantive knowledge is available thus CB-SEM can be employed to test the postulated network of relationships among the variables (i.e., test theories). One the other hand, PLS-SEM is more appropriate where theory is less developed. They are primarily used to develop theories in exploratory research.

Where meeting the ML assumptions is a challenge, PLS-SEM provides a good solution. According to Hair, Hult, Ringle, and Starstedt (2014), PLS-SEM is advantageous over the conventional SEM in situations where sample sizes are small, the data are not normally distributed, and complex models with many observed variables and relationships are estimated. Model complexity has little influence on the sample size required for PLS-SEM. In this approach to SEM, partial regression relationships are computed as opposed to estimating all the structural relationships simultaneously which is characteristic of CB-SEM. As Hair, et al., (2014, p. XII) note "PLS-SEM use has increased exponentially in a variety of disciplines with the recognition that PLS-SEM's distinctive methodological features make it a viable alternative to the more popular CB-SEM [i.e., covariance-based SEM] approach".
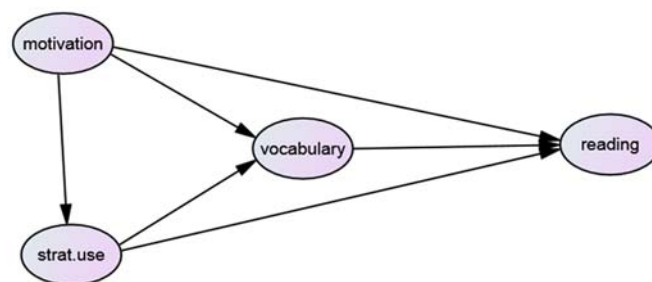
CB-SEM vs PLS-SEM

According to Chin (2010) CB-SEM is covariance based while PLS-SEM is prediction based. As a corollary of this feature, CB-SEM takes the model specified as true model and attempts to come up with path estimates that minimize the discrepancy between the model implied and sample covariance matrices. The degree of match between model implied and sample covariances is reflected in global model fit indices. Therefore, models with poor factor loadings and R-squares may show good fit indices solely on the grounds that the parameter estimates led to the highest match between the two matrices. Purely relying on model fit may result in path and factor loading estimates which are statistically significant but not substantively strong enough (Chin, 1998). Thus, Chin (1998, P. xii) suggests that "Instead, closer attention should be paid to the predictiveness of the model", rather than to overall fit.

In CB-SEM model misspecifications (e.g., adding an irrelevant item to a factor) ripple through the entire

model and affect all the path and loading estimates in the entire model. A wrongly included or exclude item under a factor affects estimates of the item loadings under another factor which might be several blocks away. However, PLS estimates are affected only by the paths and loadings in the immediate block where a given construct lies (i.e., the constructs immediately affecting or affected by the construct). Thus in the component based PLS algorithm a distinction is made between whether one wishes to explain the covariances of items under neighboring constructs or those under constructs further away.

ML estimations in CB-SEM rest on the assumptions of multivariate normal distributions and independence of observations whereas the soft distributional assumptions of the PLS-SEM render it distribution-free with no requirement of independence of observations. Consequently "traditional parametric-based techniques for significance testing/evaluation would not be appropriate" with PLS-SEM which is non-parametric and prediction oriented (Chin, 2010).

Model identification issues in CB-SEM necessitate a sample size which increases with model complexity. According to the conventional 10-time rule, 10 cases are required for any observed variable in the model which in the case of models with say, 100 items at least 1000 cases are required. However, sample size in PLS-SEM is determined by the dependent variable which has the highest number of predictors. In the model in Figure 1, imagine each of the latent variables were measured by at least 20 items which would make 80 indicators in the model. To conduct a CB-SEM a sample size of at least 800 is required whereas the same 10 time rule necessitates a sample size of only 30 cases since "reading" has the highest number of arrows

Figure 1. A hypothetical model of reading comprehension

Ravand and Baghaei: Partial Least Squares Structural Equation Modeling with R

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 3
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

directed at it, i.e., three. Therefore, PLS-SEM estimates can be obtained with much smaller sample sizes relative to model complexity.

Among less important reasons for choosing PLS over CB-SEM is model specification and model interpretation. Unlike in CB-SEM, practice of PLS-SEM does not require concerns with model identification, setting a metric for latent variables, and adequacy of the measurement scale for the estimator. In PLS-SEM, researchers simply need to specify what indicators measure each construct and how the constructs are related to each other.

Finally, PLS-SEM can handle both formative and reflective measurement models. In a reflective measurement model the construct is the cause of the indicators. Take language proficiency as a construct. The more proficient a test taker is the higher the probability of giving correct answers to language proficiency items. That is why the direction of arrows in a reflective construct are from the construct to the indicators (see Figure 2).
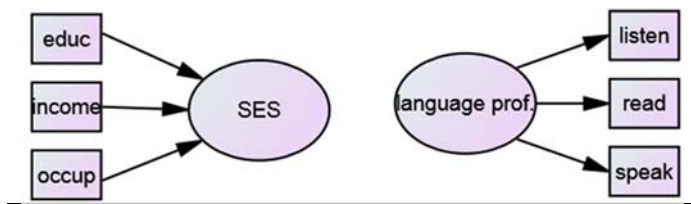
Figure 2. Examples of formative and reflective constructs. SES is a formative construct and Language Prof. is a reflective one.

On the other hand, in a formative construct the indicators cause or form the construct. Socioeconomic status (SES) is an example of a formative construct. SES is determined by education, income and occupation, among others. In a formative construct arrows go from indicators to the construct (see Figure 2). Although formative constructs can also be used with CB-SEM, "doing so requires construct specification modifications (e.g., the construct must include both formative and reflective indicators to meet identification requirements)" (Hair et al. 2014).

The merits and demerits of PLS-SEM have been heatedly debated by its proponents and critics. From the critics' camp Rönkkö and Evermann (2013) used both conceptual argument and empirical demonstration

to call into question the oft-touted capabilities of PLS-SEM such as PLS-SEM is a SEM method, can model measurement error, can be used to validate measurement models, works well with small sample sizes, provides significance tests regarding the null hypotheses of path coefficients, and is appropriate for exploratory, model building research. Henseler et al. (2014) refuted the critiques of Rönkkö and Evermann holding that their arguments stem from a flawed simulation study and misconceptions about potentials and capabilities of PLS-SEM. For more studies regarding the advantageous and disadvantageous of PLS-SEM respected readers are referred to Rigdon (2012; in press), Henseler et al. (2014), Rönkkö and Evermann (2013), and McIntosh, Edwards, and Antonakis, (2014).

As chin (2010) stated, CB-SEM and PLS-SEM should be considered as being complementary rather than competitive. Depending on the research context and the objectives, either one may be better suited. According to Henseler et al. (2014):

There is no such thing as an estimation method that is best for every model, every distribution, every set of parameter values, and every sample size. For all methods, no matter how impressive their pedigree (maximum likelihood being no exception), one can find situations where they do not work as advertised.

Hair, Ringle, and Sarstedt (2011) suggested a set of rules of thumb for choosing between PLS-SEM and CB-SEM. They suggested using PLS-SEM when the goal is theory development and theory extension which entail prediction of key constructs, when the model involves formative measures, the structural model is complex, sample size is very small, and normal distribution assumptions are violated. On the other hand CB-SEM is to be preferred in research contexts where theory testing, confirmation, and comparison is the goal hence global fit indices are required, if additional specification of error variances are needed (e.g., they need to be covariated), if the structural model includes nonrecursive relationships, and if measurement invariance of the models is to be checked.

Although PLS-SEM originated in the work of Wold (1973, 1975), it was not until 2005 that software for estimating it was made available. The late introduction of PLS software, and the dominance of CB-SEM, made PLS-SEM application lag way behind

its theoretical development. In what follows we demonstrate how to run a PLS-SEM model with R and how to evaluate and interpret the results.

Model Evaluation

As with CB-SEM, model evaluation is a two-step process: First the measurement model is evaluated and then the structural model is examined. A measurement model is evaluated in terms of the degree to which (1) the indicators are unidimensional, (2) the latent variables explain variations in the indicators, and (3) the constructs have discriminant validity.

Unidimensionality can be assessed by checking (a) Cronbach's alpha (b) composite reliability through *Dillon-Goldstein's rho*, (c) principal component analysis of each construct. Cronbach's alpha is the average correlation between the indicators of a given construct. A Cronbach's alpha index of 0.70 and higher is evidence in support of homogeneity of the indicators. Chin (1998) considers Dillon-Goldstein's rho to be a better indicator than Cronbach's alpha. As a rule of thumb Dillon-Goldstein's rho values higher than 0.70 suggest unidimensionality. Another tool for checking unidimensionality is principal component analysis of the constructs. For a block of variables to be unidimensional, their first eigenvalues should be larger than 1 while the subsequent ones should be smaller than 1.

Quality of a measurement model can also be checked by measuring how much of the variance of the indicators of a given construct is shared. The amount of shared variance is measured through establishing *convergent validity*. High factor loadings (>0.7) and average variance extracted (AVE) (>0.5) are indicators of convergent validity. AVE is the mean of *communalities* (i.e., factor loadings squared, which should be at least 0.50) of the indicators associated with any given construct. AVEs of 0.50 indicate that the construct explains at least half of the variance of its observed variables.

Discriminant validity is another tool which shows how distinct a given construct is from other constructs. Discriminant validity can be checked through (1) cross loading of the observed variables. The loadings of the indicators associated with a given construct should be higher than their loadings on any other construct. (2) Fornell-Larcker criterion which compares AVEs with the squared correlations of each construct with other constructs. For a construct to have discriminant validity, it should have an AVE larger than its highest correlation with any of the other constructs

Assessment of the Structural Model

Quality of an inner model is assessed by examining the following indices: (1) regression weights, (2) coefficient of determination (R^2), (3) redundancy index, and (4) the goodness-of-fit (Gof) statistics. Since goodness-of-fit indices of overall fit in CB-SEM are based on the discrepancy between the model implied and empirical covariance matrices, they are not suitable in PLS-SEM contexts. Fit of PLS-SEM models should depend on predictive capability of the model. In PLS-SEM statistical significance of regression coefficients is checked. R^2 values, which indicate how much of the variance in the endogenous latent variables is accounted for by their independent latent variables, have been categorized as follows:

Redundancy shows the amount of variance in an endogenous variable which is attributable to independent latent variables. No cut-off or rule of thumb has been suggested for redundancy. The higher the redundancy the higher the predictive power of the latent independent variables.

A goodness-of-fit (GoF) index has also been proposed (e.g., Tenenhaus, Amato, & Esposito Vinzi, 2004; Tenenhaus, Esposito Vizini, Chatelin, & Lauro, 2005) as a solution for the global fit of the PLS-SEM. GoF is "an index of prediction power for the entire model…. GoF values greater than 0.7 are considered as 'very good' within the PLS community" (Sanchez et al., 2015).

Previous Applications of PLS-SEM in Education

PLS-SEM has not been used extensively in education and psychology. A Scopus search with "Partial Least Square Structural Equation Modeling" as search terms resulted in 922 documents (June, 2016). Out of these, 448 were listed under 'business, management, and accounting' subject area, 266 were under 'computer science', 222 'social sciences', 140 'engineering', and 107 'decision sciences'. The 222 documents under 'social sciences' were examined. No document addressed educational and psychological constructs such as intelligence, reading comprehension, math, etc. The majority of works listed under social

Ravand and Baghaei: Partial Least Squares Structural Equation Modeling with R

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 5
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

sciences were again related to management, business, and tourism. Even those related to education addressed some management-related factors in education like student and course satisfaction and course evaluation. Below a handful of the most relevant research is reviewed.

Raffard, Bortolon, Burca, Gely-Nargeot, Capdevielle (2016) used PLS-SEM to investigate the factors that contribute to different dimensions of apathy (cognitive, emotional, and behavioral) in older adults. They found that different factors of apathy are differently and significantly related to cognitive functioning, anticipatory pleasure, sensitivity to reward, and physical functioning.

Shahijan, Rezaei, and Amin (2016) investigated the impact of perceived brand orientation, intercultural friendship, and university reputation on international students' course satisfaction and continuance behavioral intention in higher education in Malaysia using PLS-SEM. Their findings showed that perceived brand orientation and university reputation positively affect course satisfaction and course continuance behavioral intention, while intercultural friendships affects course continuance intention but not course satisfaction.

Shafaei, Nejati, and Abd Razak (2016) studied the factors that contribute to international students' psychological and sociocultural adaptation and well-being in a host country (Malaysia). They proposed a model and tested it using PLS-SEM. Their model demonstrated that longer length of stay in the host country negatively influences attachment to the home country. Furthermore, adjustment and attachment attitude positively influence psychological adaptation while only adjustment attitude positively influences sociocultural adaptation.

Hauser, Paul, and Bradley (2012) studied the association between changes to computer self-efficacy and computer anxiety and their effects on performance on computer-related tasks in online and face-to-face education. They conclude that in face-to-face education there is a need for continuous free-flowing dialog with the students to enhance learning while, logical organization is very important in online media.

Al-Azawei and Lundqvist (2015) investigated the factors that relate to student satisfaction in on-line courses in an Arabic context. Their findings showed

that learning styles and gender had no significant effect on perceived usefulness, perceived ease of use, and course satisfaction.

PLS-SEM Software Programs

Currently, there are several alternative PLS-SEM software programs such as *SPAD-PLS*, *VisualPLS* (Fu, 2006), *PLS-GUI* (Li, 2005), *PLS-Graph* (Chin, 1993–2003), *SmartPLS* (Hansmann, & Ringle, 2004), as well as the R packages *plspm* (Sanches, Trinchera, & Russolillo, 2015) and *semPLS* (Monecke, 2013). R packages *plspm* and *semPLS* like all the other R packages are free and open source. That is, the packages are not black boxes and each researcher has access to the algorithms and the codes employed behind the scenes to carry out statistical analyses and can build upon them. Access to R packages simply needs an internet connection. Statistical analyses which needs pages of codes to run, can be carried out by just a few lines of codes in R. The *plspm* package, which is applied in the current demonstration, does not allow to specify models graphically. However, it estimates confidence intervals for the parameter estimates.

Empirical Example

In this section a hypothetical data set is analyzed using plspm package (Sanches, Trinchera, & Russolillo, 2015) in R. R (R Development Core Team) is an all-in-one free statistical software which can be employed to do a wide range of statistical analyses an educational measurement practitioner might need. R and its packages can be downloaded from the Comprehensive R Archive Network's (CRAN) website at https://www.r-project.org/.Some basic information on how to download and use R is given in the Appendix.

There are four latent variables in this study: reading comprehension, motivation, strategy use, and vocabulary. The purpose of the study is to test a model of reading comprehension in relation to learners' motivation, vocabulary knowledge, and learning strategy use. PLS models are conventionally specified in two stages: First the structural or the inner model, where the relationships among the latent variables are specified, is defined. Then the *measurement* or the *outer* model, where the relationships among latent variables and their observed variables are specified, is postulated.
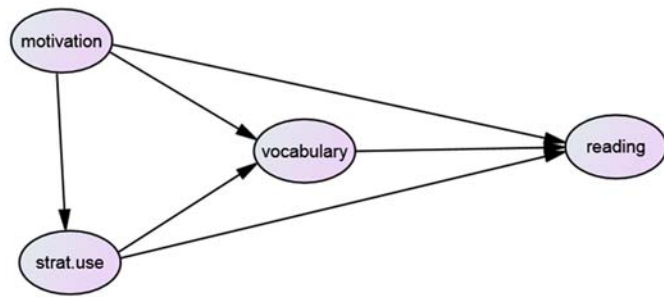
Figure 3. The postulated reading model

The structural relationships must be specified in the form of a square matrix. To define the relationships specified in Figure 3, we need a matrix with four rows and four columns (since we have four variables). In this matrix the elements in the upper triangle and the diagonal must be 0 whereas the elements in the lower triangle can be either 0 or 1 depending on the structural relationships between the constructs. As shown in Figure 4 row names and the corresponding column names are the same. For example, the first row and the first column are both 'motivation'. The matrix should be read in a "column -affecting- row" way. A number one in cell *ij* indicates that Column j affects Row *i*.

Figure 4. The square matrix to defne the model

As mentioned before, all the elements in the diagonal and the upper triangle in Figure 4 are zero. Reading in the "column-affects-row" way, the matrix indicates that 'vocabulary' (column) affects 'reading' (row), i.e., their intersection is 1. As Figure 4 shows, 'reading' is affected by all columns (variables) except itself. The zeros in the diagonal mean that the constructs cannot affect themselves and the zeros in the upper triangle indicate that no causal loops are possible in PLS-SEM (i.e., there should not be two-way arrows in the PLS-SEM).

In R four objects, each representing one of the variables, should be defined. The pattern of zeros and ones is an exact replica of the matrix in Table 2. The combine function c() is used to create the objects.

```
motivation=c(0,0,0,0)

strat.use=c(1,0,0,0)

vocabulary=c(1,1,0,0)

reading=c(1,1,1,0)
```

What we are doing is creating four vectors that will be the rows of the argument path matrix. Then we use the function `rbind()` that creates a matrix by "stacking" the row vectors. Then the vectors should be bound in a matrix using the following function:

```
inmodel=rbind(motivation,
strat.use, vocabulary, reading)
```

Now the path diagram of the structural model can be obtained by using the function `innerplot()` as follows:

```
innerplot(inmodel)
```

In the next step the outer (measurement) model is specified. In this step we have to tell the software what indicators are associated with what latent variable. To this end, we make a list containing of as many blocks of variables as we have latent constructs as follows:

```
measurmodel=list(11:20,21:30,31:3
4, 1:10)
```

In the above list four blocks of variables have been specified each representing the indicators of one of the latent constructs. It should be noted that the order of the blocks should be the same as the order of the latent variables in the matrix in Figure 4. For example indicators associated with 'motivation' are placed in Columns 11 to 20 in the data matrix, and so on.

By default, `plspm()` assumes that the measurement of the latent variables is in reflective mode, known as *mode A* in the PLS-SEM world. However, it is strongly recommended that you explicitly define a vector of modes. In the present study, for example, mode of the four constructs can be indicated through the following command:

```
mode = c("A","A","A","A") # all latent
variables are measured in a reflective way
```

Alternatively, if any of the constructs is formative (i.e., mode B), letter B is used in the above vector for that respective construct. If in the present study the first two constructs were formative, the following code should be used to reflect the mode:

Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

```
mode = c("B","B","A","A")# The first
```
two constructs are formative but the second two are reflective.

Now that the inner and outer models and the mode of the constructs have been specified, it is high time we ran the main PLS function using the following code:

```
Pls=plspm(data, innermodel, outer
model, mode, boot.val=TRUE)
```

As it was shown before, the object `mydata` was created to which the data are assigned and the `inmodel` is the object to which the inner model is assigned and the `measurmodel` is an object to which the outer model is assigned. Replacing the created objects in the above command we will have

```
Pls=plspm(mydata, inmodel,
measurmodel, mode, boot.val=TRUE)
```

To see what's included in the object Pls, simply type Pls into the editor window or R console and press 'enter'. The output shown in Table 1 is returned:

Table 1. The outputs of plspm function

Partial Least Squares Path Modeling (PLS-PM)		
	NAME	DESCRIPTION
1	outer_model	outer model
2	$inner_model	inner model
3	$path_coefs	path coefficients matrix
4	$scores	latent variable scores
5	$crossloadings	cross-loadings
6	$inner_summary	summary inner model
7	$effects	total effects
8	$unidim	unidimensionality
9	$gof	goodness-of-fit
10	$boot	bootstrap results
11	$data	data matrix

You can also use the function `'summary'`

The dollar sign ($) in R codes means that the operation specified after $ should be called within the object named before $. For example, executing the command `Pls$outer_model` will return

`Pls$unidim`. Alternatively the 'summary' function:

```
summary(Pls)
```

returns all the necessary output. Quality of a PLS path model is assessed in two stages: first the measurement model or the outer model is assessed and if its characteristics are acceptable we get to the structural model.

Results

Assessment of the measurement model

The indices related to unidimensionality of the measurement model can be recalled by executing the `Pls$unidim` command:

Pls$unidim

Table 2. Dimensionality Indices

	Mode	MVs	C.alpha	DG.rho	eig.1st	eig.2nd
motivation	A	10	0.705393	0.790631	2.824991	1.344691
strat.use	A	10	0.352403	0.581488	1.742269	1.468916
Vocabulary	A	4	0.78207	0.860024	2.427522	0.726916
Reading	A	10	0.305252	0.532348	1.711604	1.504565

Note: strat.use=strategy use; MV=manifest variable; C.alpha= Cronbach alpha;DG.rho= Dillon-Goldstein's rho; eig.= eigenvalue

According to the model evaluation criteria explained above, Cronbach alpha and Dillon-Goldstein's rho indices of 0.70 and higher support homogeneity of the indicators. As Table 2 shows both Cronbach alpha and Dillon-Goldstein's rho for both *motivation* and *vocabulary* constructs are above 0.70 whereas the other two constructs do not show acceptable values i.e., < 0.70. The unidimensionality assumption requires that the first eigenvalues of the measurement models be greater than 1 and the second eigenvalue be less than 1. As the sixth and seventh columns of Table 2 show, the first and second eigenvalues of only the *vocabulary* are acceptable.

Practical Assessment, Research, and Evaluation, Vol. 21 [2016], Art. 11

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 8
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

Table 3. Table of factor loadings (the table is truncated due to space limitations)

	Name	Block	weight	loading	communality	redundancy
1	mot1	motivation	0.134829	0.396653	1.57E-01	0.00E+00
2	mot2	motivation	0.297015	0.692382	4.79E-01	0.00E+00
3	mot3	motivation	0.092887	0.364058	1.33E-01	0.00E+00
4	mot4	motivation	0.183764	0.559904	3.13E-01	0.00E+00
5	mot5	motivation	0.202267	0.597311	3.57E-01	0.00E+00
6	mot6	motivation	0.320865	0.737344	5.44E-01	0.00E+00
7	mot7	motivation	0.126107	0.413852	1.71E-01	0.00E+00
8	mot8	motivation	0.161854	0.423113	1.79E-01	0.00E+00

Quality of a measurement model can also be inspected by checking convergent validity which shows the amount of variance the indicators have in common. High factor loadings (>0.7) and average variance extracted (AVE) (>0.5) are indications of convergent validity. As it was explained before, AVEs of 0.50 indicate that the construct explains at least half of the variance of its observed variables. As Table 3 shows, except for the indicators of the *vocabulary* and one of the indicators of the *strategy use* the entire factor loadings are below the acceptable level of 0.70. Accordingly the communalities, which are squares of the loadings, are below the acceptable 0.50 value.

Factor loadings can be recalled by executing the

pls$outer_model command.

Low communality values and as a consequence low AVEs indicate that, except for the vocabulary, the other latent variables explain less that 50% of the variability in their respective indicators. Redundancy will be referred to when evaluating the structural model. Low factor loadings and communalities as well as small AVEs (Table 5), except for the vocabulary, show lack of convergent validity in the latent variables.

Discriminant validity is another tool which shows how distinct a given construct is from other constructs. Discriminant validity can be checked by (1) the loyalty of the indicators to their respective latent variable which is reflected in the cross loading of the observed variables. The loading of the indicators associated with a given construct should be higher than their loading with any other construct. The indicators that violate this requirement are dubbed traitor indicators. (2) Any constructs should have an AVE larger than its highest correlation with any of the other constructs. As cross loadings in Table 4 show there are few traitor indicators in the model. Loyalty of the indicators supports discriminant validity of the latent variables.

Pls$crossloadings

Table 4. Table of cross loadings (the table is truncated due to space limitations)

Name	Block	motivation	strat.use	vocabulary	reading
mot1	Motivation	0.4	0.23	0.19	0.12
mot2	Motivation	0.69	0.43	0.33	0.4
mot3	motivation	0.36	0.26	0.01	0.1
mot4	motivation	0.56	0.29	0.17	0.25
mot5	motivation	0.6	0.31	0.26	0.22
mot6	motivation	0.74	0.5	0.46	0.3
mot7	motivation	0.41	0.17	0.27	0.05

However, as Table 5 shows AVE of only vocabulary is higher than its correlation with the other latent variables. Thus the second requirement of discriminant validity is not met by strat.use, motivation, and reading. Note the correlations between the latent variables (as well as all the other outputs recalled in this paper) can be generated by executing the `Pls$summary` command.

Pls$summary

Table 5. Correlations between the latent variables

	motivation	strat.use	vocabulary	AVE
motivation				0.278053
strat.use	0.588			0.167758
vocabulary	0.467	0.487		0.604537
reading	0.402	0.516	0.43	0.147677

Ravand and Baghaei: Partial Least Squares Structural Equation Modeling with R

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 9
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

Structural Model Assessment

In the next step the quality of the inner (structural) model is evaluated by executing the `Pls$inner_model` commands:

Pls$inner_model

Table 6. Structural model path coefficients

$strat.use

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.00	0.09	0.00	1.00
Motivation	0.59	0.09	6.58	0.00

$vocabulary

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.00	0.09	0.00	1.00
Motivation	0.28	0.12	2.38	0.02
strat.use	0.33	0.12	2.80	0.01

$reading

	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	0.00	0.09	0.00	1.00
motivation	0.09	0.12	0.78	0.44
strat.use	0.36	0.12	2.98	0.00
vocabulary	0.21	0.11	1.94	0.06

The first section of Table 6 shows the loading (Estimate) of 'motivation' on 'strategy use'. In the second part the loadings of 'motivation' and 'strategy use' on 'vocabulary' have been presented. And in the third section the loadings of 'motivation', 'strategy use', and 'vocabulary' on 'reading' are presented. As Table 6 shows, all the regression weights except for the effect of motivation and vocabulary on reading are statistically significant.
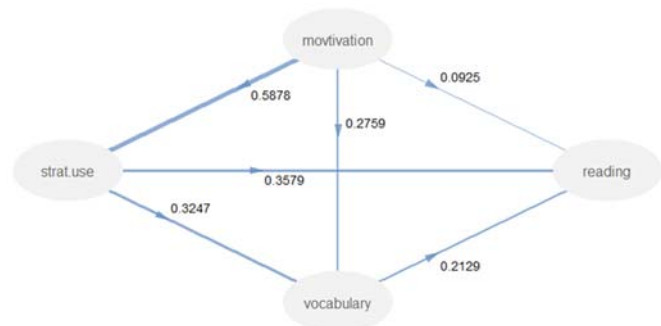
The `plot(Pls)` command returns the graphical representation of the postulated model with the regression coefficients. The loadings in Table 6 can easily be seen in Figure 5. To check the statistical significance of the loadings the table should be consulted.

Figure 5. Graphical representation of the model with the loadings

Hair et al, (2014) considered R^2 values of 0.75, 0.50, and 0.25 for the dependent variables as substantial, moderate, and weak, respectively. However, Sanchez, et al. (2015) considered R^2 values of >.60 as high, between 0.30 and 0.60 as moderate and below 0.30 as low. The R^2 values, as shown in Table 7 are low to moderate. No cut-off or rule of thumb has been suggested for redundancy. The higher the redundancy the better the predictive capability of the latent independent variables. The redundancy values for the endogenous latent variables are very low which show that the latent independent variables explain an infinitesimal amount of the variance in the endogenous latent variables.

Finally the GoF index of 0.28 for the whole model is well below the suggested cutoff of 0.70. The GOF can be recalled by executing the `Pls$gof` command.

Pls$inner_summary

Table 7. Summary of the structural model indices

	Type	R2	Block_Communality	Mean_Redundancy	AVE
motivation	Exogenous	0	0.2780534	0	0.278053
strat.use	Endogenous	0.34556	0.1677576	0.05797027	0.167758
vocabulary	Endogenous	0.2869	0.6045369	0.1734414	0.604537
reading	Endogenous	0.31342	0.1476774	0.04628506	0.147677

Practical Assessment, Research, and Evaluation, Vol. 21 [2016], Art. 11

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 10
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

In addition to the direct effects displayed in Table 6 above, indirect and total effects can also be recalled as shown in Table 8. Bollen (1987, p.40) defines direct effect as "those influences unmediated by any other variable in the model" and indirect effects as those influences "mediated by at least one intervening variable" and total effects are sums of direct and indirect effects. In Figure 5, for example, motivations affect reading both directly and indirectly through strat.use and vocabulary. The indirect and total effects of motivation on reading are calculated as follows:

Indirect effects:

P1=Motivation→ strat.use * strat.use→ reading

P2=Motivation→ strat.use * strat.use→vocabulary*strat.use-->reading

P3=Motivation→vocabulary*vocabulary-->reading

Total effect:

Total=direct path (motivation→reading) +p1+p2+p3

By executing the `Pls$effects` command, one can get the indirect and total effects:

`Pls$effects`

Table 8. Direct, indirect, and total effects

relationships	direct	indirect	Total
motivation -> strat.use	0.587843	0	0.587843
motivation -> vocabulary	0.275914	0.19088	0.466794
motivation -> reading	0.092513	0.309728	0.402241
strat.use -> vocabulary	0.324712	0	0.324712
strat.use -> reading	0.357859	0.069119	0.426978
vocabulary -> reading	0.212862	0	0.212862

Since PLS-SEM is a nonparametric statistical procedure, parametric significance tests based on normal distribution assumptions are not suitable for parameter estimates hence precision of the estimates should be checked through standard errors provided by bootstrap validation. In bootstrap N samples are drawn (with replacement) from the original data set. Therefore, N estimates for each parameter in the model is obtained and the standard deviations (standard errors) for each parameter are calculated. The size of bootstrap samples should be the same as that of the original data set. Hair et al. (2014) recommend 500 samples as a rule of thumb. For example, if our original sample has 200 valid observations, 500 samples of 200 cases are drawn randomly for the original sample then for each of the parameters in the model we obtain 500 estimates from each one of the bootstrap samples. Finally the mean of the 500 estimates and their standard deviation (standard error) is calculated. In order to obtain t-values for the parameters the original estimates of each parameter is divided by the bootstrap standard error. Bootstrap can be carried out by using the `boot.val`=TRUE in the `plspm()` function. Executing the `Pls$boot` command will return the bootstrap output shown in Table 9.

Table 9-11 show the bootstrap results for factor loadings, total effects, and direct effects, respectively. In all of these tables, the first column represents the original parameters we obtained applying the `plspm` function, the second column displays the mean of the parameters obtained from the 500 replications, and the third column displays the deviation of these estimates around their respective means (here standard errors). The last two columns display the lower percentile and upper percentiles of the 95% bootstrap confidence interval.

Ravand and Baghaei: Partial Least Squares Structural Equation Modeling with R

Practical Assessment, Research & Evaluation, Vol 21, No 11   Page 11
Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

Pls$boot
Table 9. Factor loadings and Bootstrap standard errors (the table is truncated due to space limitations)

$loadings

	Original	Mean.Boot	Std.Error	perc.025	perc.975
motivation-mot1	0.40	0.38	0.16	0.04	0.64
motivation-mot2	0.69	0.68	0.08	0.47	0.80
motivation-mot3	0.36	0.35	0.15	0.07	0.63
motivation-mot4	0.56	0.52	0.12	0.25	0.71
motivation-mot5	0.60	0.58	0.10	0.37	0.73
motivation-mot6	0.74	0.73	0.06	0.62	0.83
motivation-mot7	0.41	0.39	0.13	0.12	0.66

Table 10. Total effects and Bootstrap standard errors

$total.efs	Column1	Column2	Column3	Column4	Column5
	Original	Mean.Boot	Std.Error	perc.025	perc.975
motivation -> strat.use	0.59	0.62	0.06	0.49	0.74
motivation -> vocabulary	0.47	0.49	0.08	0.30	0.63
motivation -> reading	0.40	0.45	0.17	-0.06	0.63
strat.use -> vocabulary	0.32	0.36	0.11	0.15	0.58
strat.use -> reading	0.43	0.44	0.22	-0.24	0.65
vocabulary -> reading	0.21	0.20	0.13	-0.09	0.44

Table 11. Path coefficients and Bootstrap standard errors

$paths					
	Original	Mean.Boot	Std.Error	perc.025	perc.975
motivation-> strat.use	0.59	0.62	0.06	0.49	0.74
motivation->vocabulary	0.28	0.26	0.12	0.00	0.45
motivation->reading	0.09	0.12	0.12	-0.09	0.37
strat.use->vocabulary	0.32	0.36	0.11	0.15	0.58
strat.use->reading	0.36	0.37	0.21	-0.27	0.60
vocabulary->reading	0.21	0.20	0.13	-0.09	0.44

With the standard errors obtained from the bootstrap procedure, the significance of the PLS parameters can be tested. The parameter estimates which are at least twice their standard errors are significantly different from zero. For example as Table 11 shows, the effect of motivation on strategy use is significant at $\alpha = 0.5$ since the original estimate (0.59) is at least twice its respective standard error (t=0.59/0.06 =9.8) whereas the effect of the motivation on reading is nonsignificant (t=0.09/0.12=0.75). These results have also been

reflected in the confidence intervals associated with these parameters. The confidence intervals of the motivation-->strat.use do not contain zero hence the parameter is significantly different from zero while the confidence intervals associated with the motivation-->reading parameter contains zero, hence it is not statistically significant.

Conclusion

This paper first described the differences between PLS-SEM and CB-SEM and demonstrated the

Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

application of PLS-SEM to second language data using the freeware R. As it was shown in the Introduction section, the two SEM procedures should be viewed as complementary rather than competitive. According to Hair, Ringle, and Sarstedt (2011), in research settings where the objective is theory testing and confirmation, the CB-SEM is the right choice whereas in contexts where the aim is to make predictions and develop theories the appropriate method is the PLS-SEM. Due to its flexibility to handle both normal and nonnormal distributions, much smaller sample sizes (compared to CB-SEM), and highly complex models, PLS-SEM targets a wider range of problems than CB-SEM can in the social sciences.

In the second part of the paper R functions were provided to estimate PLS-SEM using "*plspm*" package. Step by step procedures to estimate the model were explained. A data set was analyzed and the output tables and graphics were interpreted. For beginners some introductory information on downloading and and working with R is also provided in the Appendix.

In this paper we only presented application of PLS-SEM to a simple model for those already familiar with the logic of SEM in general and PLS-SEM in particular. Having read this paper carefully, interested researchers can hopefully conduct more complicated PLS models such as multigroup PLS-SEM.

References

Al-Azawei, A., & Lundqvist, K. (2015). Learner differences in perceived satisfaction of an online learning: An extension to the technology acceptance model in an Arabic sample. *Electronic Journal of e-Learning*, 13, 408-426.

Blalock, H. M., Jr. (1986). Multiple causation, indirect measurement and generalizability in the social sciences. *Synthese*, 68, 13–36.

Chin, W.W. (2010). How to write up and report PLS analyses. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields* (pp. 655-690). Berlin: Springer.

Chin, W.W. (1993–2003). PLS Graph – Version 3.0. Soft Modeling Inc.

Chin,W.W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: results from a monte carlo simulation study and an electronic-mail emotion/adoption study. *Information System Research*, 14, 189–217.

Chin, W.W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295-358). Mahwah, NJ: Lawrence Erlbaum.

Esposito Vinzi, V., Chin, W.W., Henseler, J., & Wang, H. (2010). *Handbook of Partial Least Squares*. Berlin: Sprinegr.

Fu, J.-R. (2006). VisualPLS – Partial Least Square (PLS) Regression – An Enhanced GUI for

Lvpls (PLS 1.8 PC) Version 1.04. Taiwan, ROC: National Kaohsiung University of Applied Sciences.

Hair, F. H., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks: Sage Publications.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19, 139-151.

Hansmann, K.-W., & Ringle, C. M. (2004). *SmartPLS* Benutzerhandbuch. Hamburg: Universitaet Hamburg.

Hauser, R., Paul, R., Bradley, J. (2012). Computer self-efficacy, anxiety, and learning in online versus face to face medium. *Journal of Information Technology Education: Research*, 11, 141-154.

Henseler, J., & Sarstedt, M. (2013). Goodness-of-fit indices for partial least squares path modeling. *Computational Statistics*, 28, 565-580.

Henseler, J., Dijkstra, T. K., Sarstedt, M., Ringle, C. M., Diamantopoulos, A., Straub, D. W., . . . Calantone, R. J. (2014). Common Beliefs and Reality About PLS: Comments on Rönkkö and Evermann (2013). *Organizational Research Methods*, 17(2), 182-209. doi: 10.1177/1094428114526928

Klein, L. R. (1998). Evaluating the potential of interactive media through a new lens: search versus experience goods. *Journal of Business Research*, 41, 195–203.

Li, Y. (2005). PLS-GUI – *Graphic user interface for partial least squares (PLS-PC 1.8) – Version 2.0.1 beta*. Columbia, SC: University of South Carolina.

Loehlin, J.C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis*. Hilsdale NJ: Erlbaum.

Lohmoeller, J.-B. (1989). *Latent variable path modeling with partial least squares*. Heidelberg, Germany: Physica.

Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

McIntosh, C. N., Edwards, J. R., & Antonakis, J. (2014). Reflections on Partial Least Squares Path Modeling. *Organizational Research Methods*. doi: 10.1177/1094428114529165

Meehl, P. E. (1990). Appraising and amending theories: The strategy of lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108–141.

Monecke, (2013). semPLS: Structural Equation Modeling Using Partial Least Squares. R package version 1.0-10. https://cran.r-project.org/web/packages/semPLS/index.html .

R Core Development Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raffard, S., Bortolon, C, Burca, M., Gely-Nargeot, M.-C., & Capdevielle, D. (2016). Multidimensional model of apathy in older adults using partial least squares—path modeling. Age, 38, 1-12.

Rigdon, E. E. (2012). Rethinking Partial Least Squares Path Modeling: In Praise of Simple Methods. *Long Range Planning*, 45(5–6), 341-358. doi: http://dx.doi.org/10.1016/j.lrp.2012.09.010

Rigdon, E. E. (2014). Rethinking Partial Least Squares Path Modeling: Breaking Chains and Forging Ahead. Long Range Planning, 47(3), 161-167. doi: http://dx.doi.org/10.1016/j.lrp.2014.02.003Sanches

Rigdon, E. E. (in press). Choosing PLS path modeling as analytical method in European management research: A realist perspective. *European Management Journal*. doi: http://dx.doi.org/10.1016/j.emj.2016.05.006

Rönkkö M., & Evermann, J. (2013). A critical examination of common beliefs about partial least squares path modeling. *Organizational Research Methods*, 16(3), 425-448. doi:10.1177/1094428112474693

Trinchera, L., & Russolillo, G. (2015). plspm: Tools for Partial Least Squares Path Modeling (PLS-PM). R package version 0.4.7. https://cran.r-project.org/web/packages/plspm/index.html .

Schumacker, R.E., Lomax, R.G., (2004). *A Beginner's Guide to Structural Equation Modeling (2nd Ed)*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Shafaei, A., Nejati, M., & Abd Razak, N. (2016). Out of sight, out of mind: psychological consequences of attachment and adjustment attitude. *Studies in Higher Education*. doi: 10.1080/03075079.2016.1162780.

Shahijan, M. K., Rezaei, S., & Amin, M. (2016). International students' course satisfaction and continuance behavioral intention in higher education setting: an empirical assessment in Malaysia. Asia Pacific *Education Review*, 17, 41-62.

Temme, D., Kreis, H., & Hildebrandt, L. (2010). A comparison of current PLS path modeling software: Features, ease-of-use, and performance. In V. Esposito Vinzi, W.W. Chin, J. Henseler, & H. Wang (Eds.), Handbook of Partial Least Squares (pp.737-756). Berlin: Sprinegr.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, 48, 159-205.

Tenenhaus, M., Amato, S., & Esposito Vinzi, V. (2004). A global goodness-of- fit index for PLS structural equation modeling. In Proceedings of the XLII SIS Scientific Meeting (pp. 739-742). Padova, Italy: CLEUP.

Test & Go. (2006). *Spad Version 6.0.0*. Paris: France.

Wold, H. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (pp. 581-591). New York: John Wiley.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. G. Joreskog & H. Wold (Eds.), *Systems under indirect observations: Part II* (pp. 1-54). Amsterdam: North-Holland.

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock, A. Aganbegian, F. M. Borodkin, R. Boudon, & V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modeling* (pp. 307-357). New York: Academic Press.

Wold, H. (1973). Nonlinear iterative partial least squares (NIPALS) modelling. Some current developments. In P. R. Krishnaiah (Ed.), Proceedings of the 3rd International Symposium on Multivariate Analysis, (pp.383–407). Dayton: OH.

Wold, H. (1966). Estimation of principal component and related models by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate analysis,* (pp. 391–420). New York: Academic Press.

Appendix

Working with R

R (R Development Core Team) is an all-in-one free statistical software which can be employed to do a wide range of statistical analyses an educational measurement practitioner might need. R covers a wide range from basic statistics to more advanced techniques such as item response theory, structural equation modeling, multilevel modelling and cognitive diagnostic modeling, to name but a few. R is composed of a base program and a lot of add-on (user-contributed) packages that can be installed on the base program to perform a broad range of statistical and graphical analyses. R is syntax-based and open-source. Its language is highly flexible and customizable.

Commands can be typed into R in two ways by typing them directly into the R main window which is called *console* or indirectly into a separate window called *editor window.* A collection of commands typed into the editor window can be saved and reused for future uses. The console window automatically opens with R but to open the editor go to the **File** menu. If you want to write new commands click on **New script** but if you intend to use already-saved commands go to **Open script**. To execute commands in the console you need to simply press *enter* but in the editor put the cursor on the command and press ctrl+R

The following points should be kept in mind when working with R, especially those who are new to R:

1) R is case sensitive. For example, `Setwd()` and `setwd()` are two different functions.

2) Don't forget to use quotation marks when they are needed. For example, in the

    `install.packages("plspm")` and `setwd("c:\\temp\\R\\plspm")` functions,

    forgetting to use quotation marks would result in error.

3) Use the slash character or double backslash when specifying the path to the directory.

4) Each time a new R session starts, load the package(s) you intend to work with.

The following steps should be taken, before one can conduct any analysis in R:

1) Install the required packages,

2) Load the packages,

3) Set the directory where the data can be read from/stored into,

4) Import the data

To install a package (e.g., the plspm package), simply execute the following function:

    `install.packages("plspm", dep=TRUE)`

The argument `dep=TRUE` will also download all the packages on which the package depends. In the new dialog box, scroll down the list to select the nearest CRAN location to you and the package will be installed. Copies of R and its packages are kept up to date in different locations around the globe. To obtain graphics install "ggplot2" package using the "install.packages" function above.

Each time a new R session is started, one needs to load the packages he intends to work with. The following function can be used to load a package:

    `library(plspm)`
In general, R commands are split up into two parts as follows:

    `object <- function( arguments )`

which simply means '`object` is created from `function`'. Anything created in R is an object. Objects can be anything from single values to statistical models or the output of an analysis. For example take the following command:

Ravand & Baghaei, Partial Least Squares Structural Equation Modeling with R

```
a <- c(1,2,3)
```

The function `c` will be applied to the three values in the parentheses to make the object `a`. This function is a *concatenation* function which joins together the elements inside the parenthesis. Now we can do mathematical or statistical operations on the object 'a'. For example executing

```
a - 1
```

will yield

```
[1] 0 1 2
```

Objects are kept in memory and can be referred back for later uses. When you want to exit R, it will ask you whether you intend to save your *workspace* (the collection of objects and things you have created in a session)

The next step is to make a working directory in which you want R to read data and scripts from and store outputs, scripts, and objects into. To set the directory, use the `setwd()` command. Type in the path to the folder where you want R to read the data from and save the objects and scripts to inside the parentheses. For your convenience, you can open the folder where, for example, your data file is stored in. Click once in the address bar and the file path is selected (i.e., gets blue). Copy the path and paste it into the brackets, as follows:

```
setwd("C:\\Users\\Dropbox\\cognition")
```

Now either double the backslashes or change them into slashes (i.e., /)

After you have loaded the package, the data should be imported into R. The most convenient way to import data into R, is to save the file in the format of comma separated values (`.csv`) in Microsoft Excel or tab-delimited text (`.txt` or `.dat`).

The `foreign` package can be used to import data from different statistical packages (like SPSS or Stata) data into R. To do so, execute the following command:

```
library(foreign)
```

Depending on the format of the data `read.delim, read.csv, read.spss,or read.xlsx` functions can be used to assign the data to an object. If the data file is a `.csv` file, the following command assigns the data to the object `mydata`:

```
mydata <- read.csv("san.csv", header=TRUE) [,c(2:21)]
```

`header=TRUE` tells R to read the variable names from the first row of the data file. If the data file does not have the variable names, the argument `header=FALSE` should be used. Since item data are located in Columns 2 to 21, the brackets at the very end of the above command select all the rows and Columns 2 to 21 (within the brackets, what comes before comma refers to rows and what comes after comma refers to columns)

In a similar vein if data file is a SPSS or Excel file, the following functions can be used to import the data:

```
mydata<-read.spss("san.sav", to.data.frame=TRUE)
mydata<- read.xlsx("san.xlsx", 1)
```

in the above command we specify that the data are located in the worksheet 1 in the Excel file.

SPSS and Excel data can be imported into R by using the `Hmisc` and `xlsx` packages, respectively. First the packages must be installed and loaded as explained above.

Citation:

Corresponding Author

Ravand Hamdollah
Assistant Professor of Language Teaching
Vali-e-Asr University of Rafsanjan and University of Jiroft

ravandhamdollah [at] gmail.com