Journal of English Literature and Language
Review Article

Mediation Analysis in Linguistics: Essentials for Good Practice

Kamel Boustani

Faculty of Arts and Humanities, Manouba, Tunisia.

***Corresponding Author**: Kamel Boustani, Faculty of Arts and Humanities, Manouba, Tunisia.

Abstract:

This review article attempts to provide a comprehensive view of mediation analysis with an emphasis on its use in linguistic research. It presents an overview of the basic statistical techniques and tools necessary for the study of the mechanisms underlying the relationships between a predictor, an outcome, and an intermediate variable(s). Traditional methods of inference (e.g., the four-step approach of Baron & Kenny, the Sobel test, and the Structural Equation Modelling) and bootstrapping are described. Direct, indirect, and total effects are defined and the difference between them is clearly shown through examples. This paper, also, tries to focus on some of the most important criterion that should be considered when conducting mediation analysis in order to avoid some critical mistakes that may bias the results of the analysis (e.g., the timing criterion, confounding, sample size) and provides a short review showing the lack of mediation-based research in the field of Linguistics.

Keywords: Mediation analysis, Bootstrapping, Indirect effect, direct effect, total effect, Confounders, Sample Size, Power analysis.

Mediation analysis has become a common practice and a prominent approach used in human and social sciences to explain the different relationships between variables. It refers to a situation when the association between an independent variable and a dependent variable can be explained by their relationship to a third variable called mediator (Field, 2018, p. 652). Hayes (2018, p. 10)) notes that the value of mediation lies in the presence of a third intermediate variable that links a predictor (independent) variable and an outcome (dependent) variable. In other words, an intermediary variable (mediator) establishes an indirect pathway from an independent variable to a dependent variable. Hayes (2018, p. 3) adds that this method helps to deepen the understanding of the mechanisms through which a variable influences another one and establishes evidence or tests hypotheses about such mechanisms and boundary conditions.

Therefore, a mediation model aims at defining and describing the mechanism or the process behind an observed association of a dependent variable with an independent variable through the inclusion of a third hypothetical variable, called a mediator(s). Indeed, in addition to the direct causal relationship between the independent and the dependent variables, a mediation model implies that the independent variable affects the mediator variable which in turn influences the dependent one. Thus, the mediator variable serves to clarify the nature of the relationship between the independent and dependent variables.

As an instance of mediating mechanisms, memory processes, in cognitive psychology, mediate the transmission of information into a response. Thus, mediation reveals how a third variable affects the relation between two other variables. The methodology for evaluating mediation has made this analytical and statistical task an active research topic.

Mediation Analysis in Linguistics

Compared to its use in the field of psychology and in other domains, mediation analysis is still very little used in the field of linguistics. When querying various search engines such as Google Scholar and PsycInfo, we may notice that the use of mediation analysis in the field of linguistics is rather rare.

For instance, a query using the PsycInfo search engine (https://psycnet.apa.org/search) for peer-reviewed papers that include the word 'mediation' in the abstract section was performed in order to determine how frequent mediation analysis is used in the general fields of psychology and linguistics. This query was conducted on December 25th, 2020. Its results are presented in the following table (Table-1):

Table 1. Mediation analysis in psychology and linguistics

	PsycArticles	PsycBooks
Psychology	1,624	0
Linguistics	13	0

The above table (Table-1) showed that studies using mediation analysis in linguistics are very little (13 peer-reviewed articles found in the database of the APA Psycnet up to December 25th, 2020). By way of illustration, Francis et al. (2020) conducted a mediation analysis to investigate the role of rehearsal efficiency in word frequency and bilingual language proficiency effects. Also, Cleminshaw et al. (2020) used mediation techniques, including bootstrapping, to study social deficits in high school students with attention-deficit/hyperactivity disorder and the role of emotion dysregulation. Similarly, Crutcher & Ericsson (2000) used the controlled mediation method to examine the role of mediators in memory retrieval.

Traditional Methods of Inference Versus Bootstrapping

Traditional methods

Revisiting the historical background of statistical mediation analysis provides a wide range of approaches to test the mediation effect. A Monte Carlo study compared fourteen methods from various disciplines that were proposed to test the statistical significance of the indirect effect (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002, p. 84). One of the most important methods was proposed by Baron & Kenny (1986). It is a four-step approach in which several regression analyses are conducted, and the significance of the coefficients is examined at each step. The first step of this method consists in conducting a simple regression analysis with X predicting Y to test the direct path. In the second step, another simple regression analysis is to be carried out in order to test the effect of X on M. The third step is meant to test the significance of path b with M predicting Y. The last step is a multiple regression analysis with X and M predicting Y. If each of these tests is statistically significant, it could be concluded that there is mediation. The popularity of this approach is due to its simplicity; it is easy to understand, easy to describe and teach, and it does not require specialized software (Hayes, 2018, pp. 114-115). However, this method has been widely criticized for the fact that if, in the first regression analysis, the direct effect between X and Y is not significant, the whole analysis should be abandoned. Above all, the common criticism of this method is that it does not quantify the indirect effect and it does not require any inferential test about it (Hayes, 2018, p. 115).

Another early method for conducting mediation analysis is called the Sobel test. It is an application of the Delta Method for testing the significance of a mediation effect based on the work of Michael E. Sobel (1982). This method is based on showing that the asymptotic distribution of the indirect effect is normal and derived from the asymptotic standard error of the indirect effect. Indeed, mediation analysis using the Sobel test can examine whether an assumed mediator is significantly different from zero. It tests whether the indirect effect is significantly greater or smaller than zero. In other words, to conduct a Sobel test, the ratio of the estimate of the indirect effect is assumed to have a normal distribution under the null hypothesis that this indirect effect is zero. Then, a p-value could be calculated, based on the observed test statistic and the standard normal distribution, to be used for hypothesis testing. Alternatively, confidence intervals can be computed using the standard error and critical z-values. However, Hayes (2018, p. 521) does not recommend the use of the Sobel test for it works well only with large sample sizes.

In addition to these two methods, mediation used to be analyzed using the Structural Equation Modelling (SEM) approach which is considered to be a very powerful multivariate technique that uses conceptual models, path diagrams and systems of linked regression-style equations to capture complex relationships between observed and unobserved variables (Gunzler, Chen, Wu, & Zhang, 2013, p. 390). Several researchers and experts recommend the use of SEM approach for mediation analysis (e.g., Baron & Kenny, 1986; Frazier, Tix, & Barron, 2004) because, unlike in regression analysis, the mediator and the dependent variable can be separated from their measurement errors (Danner, Hagemann, & Fiedler, 2015, p. 463).

Bootstrapping

An alternative method of inference in mediation analysisis bootstrapping. It is a non-parametric resampling procedure used to test mediation that does not meet the assumption of normality of the sampling distribution. It is called a resampling procedure because it resamples from original data sets (Chernick, 2008, p. 1). Each time a case is drawn from the original sample, it is re-entered with the potential of being selected again as the sample size of n is constructed.

Actually, in statistical analysis, the useof bootstrapping techniques is very advantageous. The key benefit is that the normality assumptionof the sampling distribution is not required for bootstrapping. It generates an empirical approximation of the sampling distribution of the paths a and b and their product ab(these paths will be explained in section 4) by repeatedly re-sampling the data through replacement (Preacher & Hayes, 2008a, p. 26; Preacher & Hayes, 2008b, p. 880). Shrout and Bolger (2002) emphasize that the main advantage of using bootstrap methods is to detect indirect effects in a given multiple mediation model. This statement concurs Nevitt and Hancock's (2001) ideas that bootstrapping can handle the problem of non-normality in SEM by establishing empirical sampling distributions through repeatedly sampling from given original data sets.

Moreover, bootstrapping analyses are robust because they generate thousands of observations drawn from the original sample instead of the number of participants n actually included in the study. This means that marginal effects observed with normal mediation analyses generally will be significant with bootstrapping. Therefore, bootstrapping seems to reduce type II errors (type II errors happen when a null hypothesis, which is actually false, fails to be rejected. In other words, it triggers a false positive. The error rejects the alternative hypothesis, even though it does not occur due to chance) and boost power.

Direct, Indirect, and Total Effects in Mediation

In mediation analysis, it is necessary to disentangle the pathways that link an exposure (independent variable) to an outcome (dependent variable). Typically, the aim is to identify the effect of the exposure unexplained by the mediators of the model (direct effect) and the effect of the exposure that acts through a given set of mediators of interest (indirect effect).

Actually, mediation analysis is a regression analysis in which there are three key terms: direct effect, indirect effect, and total effect. The following figure (Figure-1) represents a simple mediation model in which X is the independent variable, Y is the dependent variable, and M is the mediator. The diagram shows three pathways: the pathway from the independent variable to the mediator (X→M) is labelled a, the pathway from the mediator to the dependent variable (M→Y) is labelled b, and the pathway from the independent variable to the dependent variable (X→Y) is labelled ***c'***.

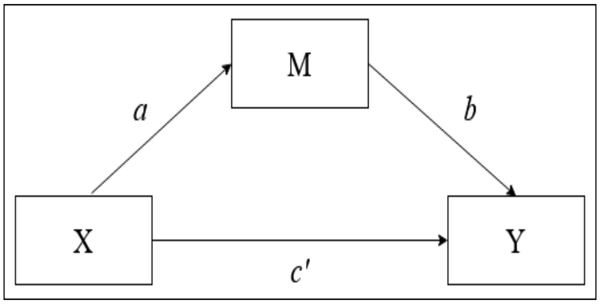

Figure 1. A mediation model with one mediator

In the diagram shown above, the direct effect, from X to Y and labelled ***c'***, estimates the extent to which the dependent variable changes when the independent variable increases by one unit and the mediator variable remains unaltered. In other words, the direct effect is the effect one could observe if the dependent variable could be changed by the independent variable without inducing a change in the mediator (Rochon, du Bois, & Lange, 2014, p. 2). Hayes (2018, p. 83) provides a basic explanation of the direct effect by claiming that two cases that differ by one unit on X but are equal on M are expected to differ by ***c'*** units on Y. The sign of ***c'*** tells whether the case one unit higher on X is estimated to be higher (***c'*** = +) or lower (***c'*** = −) on Y.

The indirect effect, labeled ab, is, on the otherhand, the product of path coefficients ***a*** and ***b***(MacKinnon, 2008, p. 50). The assumption is that these coefficients represent the impacts of X on M and M on Y, so their combination should represent the effect of X on Y through M. That is why, it is also called mediated effect. It estimates how much the dependent variable changes when the independent variable is held fixed and the mediator variable changes by the amount it would have changed, had the independent variable augmented by one unit.

In the above model (Figure-1), a quantifies how much two cases that differ by one unit on X are estimated to differ on M, with the sign determining whether the case higher on X is estimated to be higher (+) or lower (−) on M (Hayes, 2018, p. 84). For instance, if ***a*** = .85 and ***b*** = 1.3, the indirect effect will be ***ab*** = .85*1.3 = 1.105 and if ***a*** is negative and ***b*** is positive the product will be negative: e.g.,***ab*** = -.85*1.3 = -1.105.

The total effect, labelled ***c***, is the sum of the direct effect and the indirect effect ***c*** = ***c'*** + ***ab*** (VanderWeele, 2016, p. 22; Hayes, 2018, p. 85).

These effects are regression coefficients (beta coefficient = standardized coefficients) and not correlation coefficient. They do not range between -1 and +1. They are effect sizes. So, they can explain the amount of variance in the dependent variable explained by the independent variable.

In a simple linear regression, the beta coefficient represents the slope of the regression line, and in a multiple linear regression, it is the slope of the (hyper-) plane in the direction of the predictor. This means: the value of beta tells us how much the predicted value changes when the corresponding predictor is increased by 1 unit, holding all other predictors constant.

So, the standardized regression coefficients are not restricted to any range; they can take values less/more than 1. Regression actually explains that how much the value of a dependent variable will increase or decrease with the increase of the independent variable. That can be weak or moderate, but it is not shown by the range.

Basics For Good Practice

Mediational designs are central to social science and business studies (Memon, Cheah, Ramayah, Ting, & Chuah, 2018). Therefore, it is indispensable to develop a deeper scientific understanding of the processes that interfere in the relationship between the independent and dependent variables (Pieters, 2017). In the field of linguistics, for example, the processes of dealing with complex linguistic phenomena, such as affect and cognitive processes, can be refined by defining mediating factors to concentrate on particular aspects of intervention that contribute to changes in outcomes, with the possibility of discarding aspects that are less important. In fact, one practical application of mediation analysis is to evaluate the process variable that are potentially involved. This mechanism evaluation is particularly important for both theory and practice.

However, in order for the researchers not to misuse the mediation method, there are several points to be taken into consideration. First, the most critical criterion to be considered is the timing criterion of the variables. Herbert Hyman (1955), who is known as the pioneer of the original ideas of the modern mediation, claimed that only variable that are ordered in time should be treated in this type of analysis. In other words, the predictor variable should come before the mediator in time, and the mediator before the outcome variable. Second, confounding is a significant and prevalent risk to the validity of mediation research (Valente, Pelham III, Smyth, & MacKinnon, 2017). A confounder is an extraneous variable whose presence affects the relationship between the variables being studied and cause bias in the results (Pourhoseingholi, Baghestani, & Vahedi, 2012, p. 79). MacKinnon (2008, p. 7) defines confounders as variables that change the association between the independent and dependent variables because it is related to both of them. Hayes (2018, p. 122) considers confounding or spurious association as a real threat to the validity of the mediation model because they may have a significant impact on the results. Thus, researchers are said to account for their effects to avoid a false positive (Type I) error (a false conclusion that the dependent variables are in a causal relationship with the independent variable), either through experimental design and randomization before the data gathering, or through statistical analysis after the data gathering process. For this reason, a careful data treatment and statistical analysis is crucial to estimate the influence and effect of these confounding factors in the results and determine their role in the future study. Confounders are often identified through a careful investigation of the literature and through studying the possible associations of the exposure, mediator, and outcome variables with the confounders (Braga, Farrokhyar, & Bhandari, 2012, p. 133; Howards, 2018, p. 401).

Finally, another important criterion for the appropriate use of mediation is the sample size. Many researchers conduct statistical mediation without a sufficient sample size. One of the solutions for this problem is conducting power analyses which is a bit challenging for advanced methods such as mediation(Fairchild & McDaniel, 2017). Two other methods that can be applied are using data from tables in the literature or conducting a Monte Carlo study specific to the model under consideration. Researchers should reference results from current research for some types ofmediation that are specifically compatible with the model they plan to use.In other cases, a Monte Carlo analysismight be necessary to obtain adequate power estimates.to detect mediation impacts with sufficient power Fritz and MacKinnon(Fritz & MacKinnon, 2007) tables of the appropriate sample size may be referred by the researchers todetect the number of participants necessary for their proposed research.

Conclusion

The mediation methods described in this review can be effective and beneficial in the field of linguistics and its different subfields. Some possible applications of these principles and techniques include attempting to understand the various correlations between some linguistics factors, skills and subskills and the language learners' affect (anxiety, motivation, self-efficacy, etc.). However, before investigating relationships, linguistic researchers should have a good theoretical support for the existence of potentially mediating variables and should be able to manipulate theproposed mediators in an appropriate and ethical way, being conscious of confounders and other variables that can influence the relationship between the predictor and the outcome.

References

[1] Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. doi: 10.1037/0022-3514.51.6.1173

[2] Braga, L. P., Farrokhyar, F., & Bhandari, M. (2012). Confounding: What is it and how do we deal with it? *Canadian Journal of Surgery, 55*(2), 132-138. doi:10.1503/cjs.036311

[3] Chernick, M. R. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

[4] Cleminshaw, C. L., DuPaul, G. J., Kipperman, K. L., Evans, S. W., & Owens, J. S. (2020). Social deficits in high school students with attention-deficit/hyperactivity disorder and the role of emotion dysregulation. *School Psychology, 35*(4), 233-242. doi:https://doi.org/10.1037/spq0000392

[5] Crutcher, R. J., & Ericsson, K. A. (2000). The role of mediators in memory retrieval as a function of practice: Controlled mediation to direct access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1297-1317. doi:https://doi.org/10.1037/0278-7393.26.5.1297

[6] Danner, D., Hagemann, D., & Fiedler, K. (2015). Mediation analysis with structural equation models: Combining theory, design, and statistics. *European Journal of Social Psychology, 45*(4), 460-481. doi:10.1002/ejsp.2106

[7] Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics* (5th ed.). London: SAGE Publications.

[8] Francis, W. S., Arteaga, M. M., Liaño, M. K., & Taylor, R. S. (2020). Temporal dynamics of free recall: The role of rehearsal efficiency in word frequency and bilingual language proficiency effects. *Journal of Experimental Psychology: General, 149*(8), 1477-1508. doi:https://doi.org/10.1037/xge0000732

[9] Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing Moderator and Mediator Effects in Counseling Psychology Research. *Journal of Counseling Psychology, 51*(1), 115–134. doi:10.1037/0022-0167.51.1.115

[10] Fritz, M. S., & MacKinnon, D. P. (2007). Required Sample Size to Detect the Mediated Effect. *Psychological Science, 18*(3), 233–239. doi:10.1111/j.1467-9280.2007.01882.x

[11] Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry, 25*(6), 390-394. doi:10.3969/j.issn.1002-0829.2013.06.009

[12] Hayes, A. F. (2018). *Introduction to Mediation, Moderation,and Conditional Process Analysis: A Regression-Based Approach.* New York: The Guilford Press.

[13] Howards,P.P.(2018).Anoverviewofconfounding. Part 2: how toidentify it and special situations. *Acta Obstetricia et Gynecologica Scandinavica, 97*, 400–406. doi:10.1111/aogs.13293

[14] Hyman, H. (1955).*Survey design and analysis: Principles, cases and procedures*. Encino, CA: Glencoe.

[15] MacKinnon, D. P. (2008). *Introduction to Statistical Mediation Analysis.* New York: Lawrence Erlbaum Associates.

[16] MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods, 7*(1), 83-104. doi:10.1037//1082-989x.7.1.83

[17] Memon, M. A., Cheah, J.-H., Ramayah, T., Ting, H., & Chuah, F. (2018). Mediation Analysis: Issues and Recommendations. *Journal of Applied Structural Equation Modeling, 21*, i-ix. doi:10.47263/jasem.2(1)01

[18] Nevitt, J., & Hancock, G. R. (2001). Performance of Bootstrapping Approaches to Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 353-377. doi:10.1207/S15328007SEM0803_2

[19] Pieters, R. (2017). Mediation Analysis: Inferring Causal Processes in Marketing from Experiments. In P. S. Leeflang, J. E. Wieringa, T. H. Bijmolt, & K. H. Pauwels (Eds.), *Advanced Methods for Modeling Markets* (pp. 235-263). Cham: Springer International Publishing. doi:10.1007/978-3-319-53469-5_8

[20] Pourhoseingholi, M. A., Baghestani, A. R., & Vahedi, M. (2012). How to control confounding effects by statistical analysis. *Gastroenterology and Hepatology From Bed to Bench, 5*(2), 79-83.

[21] Preacher, K. J., & Hayes, A. F. (2008a). Contemporary Approaches to Assessing Mediation in Communication Research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage Sourcebook of Advanced Data Analysis Methods for Communication Research* (pp. 13-54). Los Angeles: Sage Publications.

[22] Preacher, K. J., & Hayes, A. F. (2008b). Asymptotic and resampling strategies for assessing and

comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891. doi:10.3758/brm.40.3.879

[23] Rochon, J., du Bois, A., & Lange, T. (2014). Mediation analysis of the relationship between institutional research activity and patient survival. *BMC Medical Research Methodology, 14*(9), 1-8. doi:10.1186/1471-2288-14-9

[24] Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422–445. doi:10.1037//1082-989X.7.4.422

[25] Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology, 13*, 290-312. doi:10.2307/270723

[26] VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annual Review of Public Health, 37*(1), 17-32. doi:10.1146/annurev-publhealth-032315-021402

Citation: Kamel Boustani, (2021), "Mediation Analysis in Linguistics: Essentials for Good Practice", Journal of English Literature and Language; 2(1): 1-7.

DOI: 10.31829/2767-2964/jell2021-2(1)-106