CHAPTER N

**Mediation, Moderation, & Interaction**

*Definitions, Discrimination & (Some) Means of Testing*

INTRODUCTION

In 1986 Baron and Kenny set out to clarify the terms “Mediation” and “Moderation” as used in the social sciences (with the origins of each described by Roe, 2012). Twenty six years later, the seminal paper that this collaboration resulted in (Baron & Kenny, 1986) has been cited over 35,000 times (35,672 via Google Scholar as of 09/01/2013). However, despite this extensive record of citation, uncertainty continues to surround the use of these terms in social science research and they have received relatively little attention in specifically educational research (cf. Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001). Partly in response to this uncertainty, and partly in response to advances made in the application of more complex statistical analyses in educational research (e.g. Creemers, Kyriakides, & Sammons, 2010; Goldstein, 2003; Luyten & Sammons, 2012; Tatsuoka, 1973), this chapter is made-up of four sections which together provide the quantitative educational researcher with an up to date understanding of these terms as well as examples of their current implementation to test theoretical models and address notions of causality. These four sections are titled:

1. Unambiguous Definitions
2. Discriminating Mediation, Moderation, and Interaction
3. Some means of testing Mediation and Moderation
4. Testing Moderation: An example through three equivalent statistical

analyses

Together, the first two sections of this chapter present simple, clear definitions that distinguish “Mediation”, “Moderation”, and “Interaction” both from each other as well as from a number of other commonly-used terms. Section 3 then presents a number of statistical methods by which these terms can be statistically operationalised. This third section pays particular attention to Moderation as the statistical methods associated with it (in comparison to Mediation) are particularly varied and numerous. The final section of this paper (Section 4) then builds upon the focus on Moderation within Section 3 by presenting an example Moderation from educational research conducted within the early years (for children under age 5 years) which is then statistically operationalised and tested by three equivalent parallel analyses.

1. Unambiguous Definitions

*Mediation*

This is a trivariate one-tailed hypothesis concerning mechanisms of effect. A pre-established causal relationship between two variables is theorised to exist due to an intermediate third variable (see Figure 1). While the additional (third) variable that is hypothesised to have this effect is known as a “*mediator*” it is also sometimes referred to as an “*intermediate variable*” (Kraemer *et al*., 2001), or “*explanatory link*” (Rose, Holmbeck, Coakley, & Franks, 2004). Further, mediators have “*mediating effects*” which are otherwise labelled “*indirect effects*”, “*surrogate effects*”, “*intermediate effects*” and/or “*intervening effects*” (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Wu & Zumbo, 2007).

*Moderation*

This is also a trivariate one-tailed hypothesis – but one that is quite different from mediation and with a completely separate historical origin (Roe, 2012). Hypotheses of moderation ask, “Under what conditions/for whom/when is a pre-established causal relationship observable?” (cf. mechanisms in mediation). The presence of a third “*moderator*” variable is also termed an *“effect-modifier”* (Hinshaw, 2002) and/or a “*causal interaction effect*” (Wu & Zumbo, 2007). This last alternative name is also important as it evidences the close association (and therefore also sometimes confusion) between the terms “Moderation” and “Statistical Interaction” (an explanation for the alternative name of “casual interaction effect” is given in Section 3). The origins of this association go back to the first reported use of “Moderation” – commonly cited as Saunders (1955) – in which this term was adopted as a synonym for what quantitative researchers now refer to as a “*(Statistical) Interaction Effect*” (again, see Section 3). This change in meaning over the past 58 years and the close relationship that today’s definition of “Moderation” has to “(Statistical) Interaction” (see below) is just one reason why confusion continues with the use of these terms.

*(Statistical) Interaction*

This is a two-tailed hypothesis implying that two or more concepts, “*work together*” or, “*have a combined effect*” in eliciting a third (for example in: Kraemer *et al*., 2001; Talamini *et al*., 2002) which should in no way be mistakenly confused with the concept of behavioural or psychological or gene-environment interactions (e.g. Rutter & Silberg, 2002). One of the common points of difficulty (explored further in Section 3) is that Moderation is a hypothesis that is often answered by the specification of a ‘Statistical Interaction’ which is then, in-turn, commonly tested with statistical artefacts known as *“Statistical Interaction Effects/Terms”*.

*C:\Users\JamesH\Desktop\1st author paper collabs\Hall and Sammons - mod med and inter\figures for book chapter submission\Figure1.eps*

*Figure 1. Graphical Illustration of the hypotheses of Mediation and Moderation*

2. Discriminating Mediation, Moderation, and Interaction

*Mediation ≠ Moderation*

Although Mediation and Moderation are distinct trivariate research hypotheses, confusion continues not only over their distinction, but also over which is the more appropriate for any given research project as well as how these hypotheses can be combined. The first of these difficulties (distinguishing Mediation and Moderation) continues partly due to the simple similarity of the two words, partly due to their changing definitions over time, and partly due to the similar purposes for which both are used in research. Considering this third point in more detail: Mediation and Moderation are both, *“theories for refining and understanding a causal relationship”* (Wu & Zumbo, 2007) and both are unidirectional (i.e. “A and B affect C” rather than “there is an association between A, B, and C”) trivariate hypotheses. The problems that arise over the application of these distinct hypotheses is also evident in the continuing confusion concerning the term “*Indirect Effect*” which although having a specific meaning encompassing Mediation (see Preacher & Hayes, 2004) also has an additional and more intuitive meaning: “*any and all effects other than those direct*”. This additional understanding of the term “Indirect effect” has led to its use in reference to Statistical Interaction and thereby also Moderation. The paper by Goodnight, Bates, Staples, Petit, and Dodge (2007) provides an example of this more intuitive usage of the term, although for clarity we recommend such usage should be avoided,

…However, in addition to direct main-effects-type links between temperament and behavior problems, there are also more indirect, interaction-effect-type links involving temperament...

With this background of confusion over the meaning and usage, the terms “Mediation” and “Moderation” have continued to be discussed long after the paper of Baron and Kenny (1986). Table 1 provides an overview of a selection of five journal articles in different fields since the turn of the millennium that have all aimed to provide clarifying guidelines. The problems researchers continue to encounter with these terms is evidenced in the inconsistent guidelines across these papers. A casual examination of Table 1 also reveals that none of the articles originate from the field of Educational Research and, to the best of our knowledge prior to this chapter; Educational Researchers have never had tailor-made guidance written for them on the issues that surround Mediation and Moderation.

*Table 1. A selection of past guidelines (since 2000) for distinguishing Mediation from Moderation*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Authors:**  (Kraemer, *et al*., 2001) | (Hinshaw, 2002) | (Nicholson, Hursey, & Nash, 2005) | (Essex, *et al*., 2006) | Wu & Zumbo (2007) |
| **Journal:**  *American Journal of Psychiatry* | *Development and Psychopathology* | *Headache* | *Archives of General Psychiatry* | *Social Indicators Research* |
| **Mediation:** | | | | |
| That being mediated has temporal precedence | | | | |
| *Mediator and that mediated are correlated* | | | | |
| Either co-domination of mediated and mediator (partial) OR |  |  | Either co-domination of mediated and mediator (partial) OR | Answers, "how" and "why" |
| *Mediator dominates that mediated (total)* |  |  | *Mediator dominates that mediated (total)* | *Mediator is a state* |
|  |  |  |  | Mediator is observed or manipulated |
| **Moderation:** | | | | |
| Moderator has temporal precedence | | | | |
| *Moderator and that moderated are uncorrelated* | | | | |
| Co-domination of moderated and moderator |  |  |  | Answers, "for whom" and "when" |
|  |  |  |  | *Moderator is a trait* |
|  |  |  |  | Moderator is observed |

Of the guidelines distinguishing Mediation from Moderation that are presented in Table 1, only two are consistent across all the articles:

1. The varying importance of temporal precedence (e.g. Cole & Maxwell, 2003)
2. The varying importance of which measures should/should not be significantly correlated with one another.

The first of these is particularly paramount given that both Mediation and Moderation are viewed as causal unidirectional hypotheses of effect. As a result, they require appropriate quantitative data to be tested: that which is appropriate for testing *any* unidirectional hypotheses. This is a condition of gathered quantitative data that is most commonly resolved by collecting data with a temporal element (i.e. data that is *longitudinal* in the case of correlational/survey research or *repeated-measures* in the case of experimental designs). Having data with the correct clear temporal precedence (establishing ‘causal priority’; Preacher & Hayes, 2004) is perhaps the most important precondition that researchers can and should establish for both Mediation and Moderation.

*Real-World Ambiguities.* Unfortunately, even when educational researchers hold clear unambiguous definitions of Mediation and Moderation there still remain real-world occasions in which the appropriateness of one over the other is ambiguous. Within developmental science (a catch-all label that includes much quantitative educational research), this can often be attributed to the time-frame under consideration. For example, it is often possible for the same set of measures to be related first as a mechanism (mediation) but then later as a conditional effect (moderation). The paper by Masten (2007) provides an example of this. The relationship between background adversity, an individual’s stress-regulators, and their subsequent stress-response begins with stress regulators being shaped by adversity as they develop. However, once stress-regulators are developed, their relationship with adversity changes: stress-regulators are now deemed to operate by altering the stress-response to adversity. Thus in the first period, a *mechanism* (mediation) is at work while in the second, a *condition*al effect (moderation) comes into evidence (see Figure 2).

*C:\Users\JamesH\Desktop\1st author paper collabs\Hall and Sammons - mod med and inter\figures for book chapter submission\Figure2.eps*

*Figure 2. The plausibly of Mediation and Moderation as appropriate hypotheses varying by the time-frame under investigation (adapted from Masten, 2007)*

Another common difficulty that researchers can face when determining whether it is more appropriate to specify a hypothesis of Mediation or Moderation is that these hypotheses can also be combined. Again Masten (2001) provides an example - this time of, *“...a risk-activated moderator analogous to an automobile airbag or immune system response”*. Figure 3 illustrates this effect with generalised labels. For the educational researcher in particular, this is also an excellent description of the ideal functioning of social interventions such as Head Start in the USA (see Currie and Thomas, 1995) and Sure Start in the UK (see Glass, 1999). At the same time, although this example has obvious application in the real-world it also contravenes the only two consistent guidelines about when to hypothesise Moderation that are shown in Table 1.

*C:\Users\JamesH\Desktop\1st author paper collabs\Hall and Sammons - mod med and inter\figures for book chapter submission\Figure3.eps*

*Figure 3. An example moderation that contravenes common guidelines but which has real-world application (adapted from the “risk-activated moderator” of Masten, 2001)*

The “*risk-activated moderator*” of Masten (2001; Figure 3) is just one example of how Mediation and Moderation may be integrated as hypotheses. Two more examples of these hypotheses in combination are *“Moderated-mediation”* and *“Mediated-moderation*” (see Figure 4). Both hypotheses concern causal relationships hypothesised between at least four measures [W, X, Y, Z] and both postulate *conditional* (moderated) *mechanisms* of effect (mediation) whereby X affects Y. Further in-depth description and discussion of these combinations of Mediation and Moderation can be found in Muller, Judd, and Yzerbyt (2005), Wu and Zumbo (2007), Preacher, Rucker, and Hayes (2007), and Edwards and Lambert (2007). These papers also outline the various approaches for the statistical testing of these hypotheses.

*C:\Users\JamesH\Desktop\1st author paper collabs\Hall and Sammons - mod med and inter\figures for book chapter submission\Figure4.eps*

*Figure 4. An illustration of the hypotheses of Mediated-Moderation and Moderated-Mediation (adapted from Wu & Zumbo, 2007)*

*Moderation ≠ (Statistical) Interaction*

Just as hypotheses of Moderation are often confused with Mediation so too is Moderation often confused (or viewed as synonymous with) Statistical Interaction. This is a problem that is at-least partly due to the overlap between the two concepts, one where Moderation can be viewed as a more restricted version of Statistical Interaction as evidenced by the alternative name for Moderation given by Wu and Zumbo (2007): “*Causal Interaction*”. The relationship between the concepts of Statistical Interaction and Moderation can be understood as the difference between a two-tailed hypothesis and a more restrictive one-tailed hypothesis. Thus, although the statistical methods that are used to test hypotheses of (Statistical) Interaction can also be applied to hypotheses of Moderation, to conclude Moderation from these methods necessitates relying heavily upon pre-existing knowledge, be this from past research findings, broader substantivetheories, or from other top-down sources of knowledge (e.g. Nicholson, Hursey, & Nash, 2005). When there is a lack of sufficient pre-existing knowledge to justify putting forward a hypothesis of (one-tailed) Moderation at the beginning of a research project, subsequent evidence of a (two-tailed) Statistical Interaction should not over-interpreted as inferring Moderation (as in Rutter & Silberg, 2002; Kraemer *et al*., 2001) . For example, finding evidence that educational outcomes are significantly related to the interaction of parental background and various educational factors that a child experiences (thus: attainment=background + ed.factor + background *x* ed.factor) should not be over-interpreted to conclude that education can alter the effects of parental background unless there is additional top-down information to warrant this (e.g. Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Hall, *et al*., 2009; NICHD, 2000).

*(Statistical) Interaction ≠ Statistical Interaction Terms*

While Statistical Interaction is a two-tailed hypothesis that two or more concepts “*work together*”/“*have a combined effect*” upon a third, Statistical Interaction Terms are two-tailed statistical artefacts (defined as the product of two variables) and are often specified to test these hypotheses - commonly within regression- (of bivariate form: Y=b0+b1X+b2Z+B3X.Z+e) and ANOVA-based statistical analyses.

The relationship between Moderation, Statistical Interaction, and Statistical Interaction Terms takes the following form: *Moderation* is a more restricted one-tailed alternative to the two-tailed hypotheses of *Statistical Interaction* although both are often tested through the specification of *Statistical Interaction Terms.*

3. Some means of testing Mediation and Moderation (theory in practice)

*Mediation*

Although MacKinnon and colleagues (2002) discuss fourteen statistical methods to test hypotheses of Mediation, here we note only the four main methods and direct readers to Hayes (2009) for a fuller while also contemporary discussion of Mediation as well as the methods that are available for its testing.

1. The ’C*ausal Steps Approach’* of Baron and Kenny (1986). This is a technique that has been strongly criticized as having the least statistical power to accurately detect Mediation Effects (Fritz & MacKinnon, 2007; Hayes, 2009)
2. The *Sobel Test* (Sobel, 1982) is a more formal test of mediation compared to the Causal Steps Approach. Multiple regression analyses are conducted and the results of each are combined (see Preacher & Hayes, 2004). Various macros and online calculators are available for this additional step (for example from: http://www.danielsoper.com)
3. *Bootstrapping*. One of the problems with the Sobel Test is that it assumes normality in the distribution of variables which limits its appropriate application. One alternative that does not make this assumption is to conduct statistical boostrapping to estimate Mediation Effects. Not only is this non-parametric technique applicable with non-normally distributed variables, it is also retains its reliability with lower sample sizes compared to the Sobel Test (for more detail see Preacher & Hayes, 2004).
4. Statistical *Path Analysis* (often within “Structural Equation Modelling”, SEM) commonly incorporates the above *Bootstrapping* approach within a broader statistical modelling framework that Reynolds and Ou (2003) note as an especially suitable technique for*, “theory driven tests of hypotheses of causal mediation”* (p.451, Reynolds & Ou, 2003). A good overview is provided by Tatsuoka (1973) who documents both the historical origins of path analysis and provides an account of its initial take-up by educational researchers.

*Moderation*

Compared to the methods available for testing a hypothesis of Mediation, the options available to researchers interested in Moderation are both more numerous and more complex (with this at least partly attributable to the changing definition of Moderation over time and partly due to its conceptual relationship with Statistical Interaction). Back in 1986 Baron and Kenny discussed the statistical methods suitable for testing hypotheses of moderation (statistical interaction terms within either regressions or ANOVAs) and presented detailed guidelines for deciding the appropriateness of one over the other. For this, Baron and Kenny emphasised that the suitability of a method depended upon the level of measurement by which each of the three concepts featured in the Moderation were measured, be this continuous or categorical. Fifteen years later however and Kraemer and colleagues (2001) noted a “*struggle*” existed between two statistical approaches used for testing hypotheses of moderation: 1) Sub-group comparisons (that commonly dichotomise continuous moderator variables), 2) Statistical Interaction Terms. For the educational researcher in particular however, there is at least one additional statistical technique not covered by either Baron and Kenny (1986) or Kraemer and colleagues (2001) and which also directly tests hypotheses of Moderation, has nothing to do with Statistical Interaction Terms (in ANOVAs or regressions), and for which quantitative educational data is frequently suitable: *Random Slope Effects*. These are typically examined in multilevel models that explore the hierarchical structure of nested data in social or educational contexts where students are clustered in classes, themselves clustered in schools, in turn clustered in neighborhoods etc (see Goldstein, 2003; Luyten & Sammons, 2010). Here examples of hypotheses that may be tested include that the shape of relationships between prior attainment and later attainment (the slope) may differ between higher level units (for example classes or schools) and also for different groups of students within different schools (for example by SES or gender)

An in-depth discussion of the three methods (sub-group comparisons, statistical interaction terms, random slope effects) that are particularly suitable for educational researchers interested in testing hypotheses of moderation follows below. First however, it is worth mentioning that these options may be grouped in two different ways: 1) Explicit versus Implicit Tests, and 2) Variable-based versus Person‑based. The Explicit/Implicit distinction refers to whether a technique is, or by contrast is not, a literal and direct test of the trivariable causal hypothesis of Moderation that is illustrated in Figure 1 (random slope effects) or whether a conclusion of Moderation is instead only inferred from a Statistical Interaction (sub-group comparisons, statistical interaction terms). The Variable/Person-based distinction refers to whether a method emphasises a pattern of statistical relationship between *variables* (as in random slope effects, statistical interaction terms) or statistical differences between *units of analysis* (commonly people; as in sub-group comparisons).

1. *Sub-group comparisons* (indirect person-based test of moderation). Here, evidence of moderation is obtained by establishing that a bivariate relationship is significantly different between two or more groupings of the unit of analysis (be these people, schools etc). If the moderator variable was originally measured with a continuous variable, then an intermediate, though strongly criticised (Frazier, Barron, & Tix, 2004; MacCallum, Zhang, Preacher, & Rucker, 2002; McClelland & Judd, 1993), step is necessary: sub-group creation through dichotomisation/ categorisation.
2. *Statistical Interaction Terms* (indirect variable-based test of moderation). This is a multi-stage procedure that only actually tests the existence of a combined working-together of two or more variables as they jointly impact another. It is then up to the researcher to interpret whether this also constitutes evidence of moderation (see Wu and Zumbo, 2007). As an act of inductive reasoning, this additional step should be informed by broader conceptual understanding such as the findings from previous research. This means of testing explains the alternative name for Moderation as a, “*Causal Interaction Effect*”: A non-causal bi-directional relationship is established (the Interaction Effect) before post-hoc reasoning is undertaken to establish a causal relationship from this. The procedure for testing a Statistical Interaction Term in an OLS regression is as follows:
   1. Mean-centre your predictor [X] and moderator [Z] variables
   2. Construct a new ‘interaction’ variable of the form: predictor x moderator [XZ]
   3. Use this variable as a predictor of your outcome [Y] along with the original variables [X, Z] in a regression equation of the form: Y=X+Z+XZ [+e]
   4. Interpret only the *un*standardised regression co-efficient from the [XZ] statistical interaction term.
   5. Plot any significant Statistical Interaction Term to aide interpretation
3. *Random Slope Effects* (direct variable-based test of moderation). Unlike Sub-group Comparisons and the use of Statistical Interaction Terms, Random Slope Effects test a hypothesis of Moderation directly, not via the intermediate step of first establishing a Statistical Interaction. Random Slope Effects refer to when a statistical regression relationship (the Slope, [s]) between two variables [X, Y] is allowed to vary as a function of a third [Z]. Unfortunately, this most direct means of testing a hypothesis of Moderation is also the most restricted in terms of the requirements it imposes on quantitative data. Random Slope Effects require clustered or nested data (as noted above) and therefore multilevel (hierarchical linear) statistical modelling techniques. On top of this, for a Random Slope Effect to test a hypothesis of Moderation a very specific set of relationships must to be specified between variables: The Moderating variable [Z] must be at the *between* level (level 2) while the Moderated relationship [Y on X] must be at the *within* level (level 1). Fortunately, the requirement for nested quantitative data is one that educational research frequently meets due to the nested nature of educational systems (e.g. children within classes within schools within districts/neighbourhoods).

4. Testing Moderation: An example through three equivalent statistical analyses

The final section of this paper presents a worked example of some of the main issues so far discussed. An educational research question is expressed as a hypothesis of Moderation and this is then tested with the three statistical approaches discussed above in Section 3: A Sub-group Comparison, specification and testing of a Statistical Interaction Term, and the testing of a Random Slope Effect.

*Theoretical Background*

A mother’s age at the birth of her child is known to significantly impact this child’s cognitive development: *Children of younger mothers are likely to demonstrate poorer cognitive development* (Borkowski, *et al*., 1992; Fergusson & Lynskey, 1993). However, higher ‘quality’ (Currie, 2001) preschool has been found to partial ‘protect’ (Rose, *et al*., 2004) children from such adverse outcomes (Burchinal, Peisner-Feinberg, Bryant, & Clifford, 2000; Hall, *et al*., 2009; NICHD, 2000). This set of relationships can be expressed as a hypothesis of Moderation: Attendance at a preschool of higher quality may *moderate* the relationship between a mother’s age (at child-birth) and her child’s subsequent cognitive development.

*Method*

To test the hypothesis of Moderation suggested above, a secondary analysis of the Effective Preschool, Primary, and Secondary Education (EPPSE; Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2010, 2012) dataset was undertaken. Adopting a longitudinal research design, EPPSE was the first large scale British research project to examine the quality and effectiveness of various programmes of pre-, primary, and secondary schools as predictors of the development and educational attainment of over 3,000 British children from 3 years of age to adulthood.

*Participants.* 2857 participating families with children in attendance at n=141 preschools (for at least 10 weeks already) were recruited when these children were of mean age 36 months. This recruitment of families from preschools ensured that a sufficient level of nesting of data was achieved (families within preschools) such that preschool effects on child outcomes could be reliably estimated (e.g. Goldstein, 1987, 2003).

*Measures.* For this example, the outcome measure [Y] was each child’s General Cognitive Ability (GCA) as measured by the British Ability Scales (Elliot, NFER-NELSON, Smith, & McCulloch, 1996) and as assessed at mean child age 58 months (n=2574; mean=96.73; Standard Deviation, SD=14.51). The predictors of GCA at mean child age 58 months were:

* GCA at 36 months (n=2764; mean=91.36; SD=13.90)
* Mother-age at child-birth[X] assessed at parental interview at enrolment (n=2779) with a six category ordinal scale (1=*16-20*, 2=*21-25*, 3=*26-35*, 4=*36-45*, 5=*46-55*, 6=*56-65*) that is here treated as continuous for solely pedagogical purposes (thus: mean=3.16; SD=0.66)
* The hypothesised moderator [Z]: The overall ‘quality’ of the preschool that each child attended as measured by the Early Childhood Environmental Rating Scale (ECERS-R; Harms, Clifford, & Cryer, 1998; preschool n=141; child n=2857; child mean=4.47; SD=1.00)

With reference to the guidelines of Table 1, it should be noted that the hypothesised moderator, preschool quality, was uncorrelated with the variable whose effect quality was hypothesised to moderate (mother’s age). For more details on these measurements see Sylva and colleagues (2010).

*Analytic Techniques.* Each of the three statistical techniques for testing a hypothesis of Moderation that were discussed in Section 3 (Sub-group Comparisons, Statistical Interaction Terms, Random Slope Effects) were conducted within the statistical framework of Multilevel Structural Equation Modelling (SEM) using Version 6 of the Mplus Software (Muthén & Muthén, 2010). Version 6 of the Mplus Software estimated missing data using maximum likelihood procedures as an integral part of all three analyses (Muthén & Muthén, *ibid*). As the following results serve only as an example of the methods discussed above, we do *not* report the results in full as we would in a purely substantive piece of work as a detailed substantive interpretation is not the aim (full results are of course available from the authors).

*Results*

*Sub-group Comparisons.* Comparisons between sub-groups based on the quality of preschools were made possible through the specification of a “multi-level mixture model”. Given that preschool quality was originally measured on a continuous scale, the specification of sub-groups necessitated an initial step of dichotomisation. A mean ± 1 standard deviation dichotomisation strategy was used to form two groups of n=623 and n=461 children who had attended n=55 ‘low’ and ‘high’ quality preschools respectively (the remaining and excluded n=1773 children attended n=86 preschools where quality was within the mean±1SD range). The following effects of mother-age on GCA at 58 months were found:

* In the ‘Low’ preschool quality group: Older mothers had children who demonstrated significantly higher GCA (standardised regression coefficient, β=0.10, p=0.001)
* In the ‘High’ preschool quality group: There was no significant relationship between mother age and child GCA (β=-0.01, p=0.85)
* Further, the relationship between mother age and child GCA was significantly higher in the ‘Low’ quality preschool group than it was in the ‘High’ quality group (β=0.10 vs. β=-0.01; t1080=2.66, p<0.01)

In conclusion, a differential impact of mother’s age upon GCA at 58 months was found in low versus high quality preschools. For children attending ‘high’ quality preschools, the children of younger mothers had (on average) indistinguishable levels of GCA compared to children of older mothers: This was not so for children attending ‘low’ quality preschools

MPLUS SUBGROUP COMPARISON (VIA DICHOTOMISATION) SYNTAX:

MISSING ARE ALL (-999999);

idvariable = childid;

CLUSTER = centreid;

WITHIN = bgcam q53am;

CENTERING = GRANDMEAN (bgcam, q53am);

CLASSES = group (2);

KNOWNCLASS = group (group1=0 group1=1);

**DEFINE:**

IF (ecers\_r LE 3.4690) THEN group1=0;

IF (ecers\_r GE 5.4691) THEN group1=1;

**ANALYSIS:**

TYPE = MIXTURE TWOLEVEL;

ALGORITHM = INTEGRATION;

**MODEL:**

%WITHIN%

%OVERALL%

rgcam on bgcam;

rgcam on q53am;

%group#1%

rgcam on bgcam;

rgcam on q53am;

%group#2%

rgcam on bgcam;

rgcam on q53am;

*Statistical Interaction Terms.* A “multi-level path model” was specified in which the Statistical Interaction Term introduced in Section 3 (Y=X+Z+XZ) was specified at the preschool (between) level. The following effects of mother-age on GCA at 58 months were found (bearing in mind that standardised results were unavailable for this model):

* There was a positive and statistically significant effect of mother-age on GCA at 58 months (unstandardised regression coefficient, b=4.92, p<0.001)
* There was also a positive effect of preschool quality on GCA at 58 months although this did not quite reach the 95% confidence level (b=1.780, p=0.064)
* There was also a statistically significant negative effect from the Statistical Interaction Term *mother-age x preschool quality* (b=-0.685, p=0.020)

In conclusion: mother’s age had a decreasing effect on children’s GCA at 58 months as the preschools that these children attended increased in quality. The children of younger mothers had (on average) lower GCA at 58 months but this was less apparent when these children had attended higher quality preschools. In other words, this analysis and its results lead to the same substantive conclusion as that returned from the Sub-group Comparisons.

MPLUS INTERACTION TERM SYNTAX:

MISSING ARE ALL (-999999);

idvariable = childid;

CLUSTER = centreid;

BETWEEN = ecers\_r;

WITHIN = q53am bgcam qualage;

CENTERING = GRANDMEAN (bgcam, q53am, qualage);

DEFINE: qualage = ecers\_r\*q53am;

**ANALYSIS:**

TYPE = RANDOM TWOLEVEL;

ALGORITHM = INTEGRATION;

**MODEL:**

%WITHIN%

rgcam on bgcam q53am qualage;

%BETWEEN%

rgcam on ecers\_r ;

*Random Slope Effects.* Once again, a “multilevel path model” was specified, but this time also featuring “random effects”. In this analysis only one random effect was estimated: the statistical regression slope (s) between mother’s age and child GCA at 58 months was allowed to vary between children and this variation was specified to depend upon the quality of the preschools that children attended. As with the estimation of the Statistical Interaction Term, the specification of a Random Slope Effect meant standardised results were again unavailable. The following effects of mother-age on GCA at 58 months were found:

* There was a positive and statistically significant effect of mother-age on GCA at 58 months (unstandardised regression coefficient, b=4.54, p=0.001)
* There was no effect of preschool quality on GCA at 58 months (b=-0.44, p=0.208)
* The significant relationship between mother-age on child GCA at 58 months was significantly diminished by increasing preschool quality (b=-0.63, p=0.031)

In conclusion, mother’s age had a smaller effect on child GCA at 58 months when these children were in attendance at higher quality preschools. Once again, this conclusion is essentially the same as that drawn from the Sub-group Comparisons and the specification/testing of the Statistical Interaction Term discussed above.

MPLUS RANDOM SLOPES SYNTAX:

MISSING ARE ALL (-999999);

idvariable = childid;

CLUSTER = centreid;

BETWEEN = ecers\_r;

WITHIN = q53am bgcam;

CENTERING = GRANDMEAN (bgcam, q53am);

**ANALYSIS:**

TYPE = RANDOM TWOLEVEL;

ALGORITHM = INTEGRATION;

**MODEL:**

%WITHIN%

rgcam on bgcam;

s | rgcam on q53am;

%BETWEEN%

s on ecers\_r;

rgcam on ecers\_r;

*Discussion*

Although all three methods led to the same substantive conclusion, the robustness of the relationship between this conclusion and the various statistical results/evidence varied. For example, the use of Sub-group Comparisons meant that no estimation was possible of any direct effect from preschool quality on child GCA at 58 months. Furthermore, although the specification of a Statistical Interaction Terms did include this estimate, the operationalisation of the hypothesised moderation was weaker than with the Sub-group Comparisons. This was because the specification of the Statistical Interaction Term was equivalent to using a two-tailed statistical technique to test a one-tailed hypothesis. It was only through the use of a Random Slope Effect that an appropriate one-tailed statistical test was carried out for a one-tailed hypothesis while the conducted analysis also fully estimated the effect of preschool quality upon child GCA at 58 months (leaving aside the potential problems that come through dichotomising moderators as here in the Sub-group Comparisons). The three worked examples show that it is important for educational researchers to specify their causal hypotheses carefully and to be aware that the robustness of the results and the conclusions drawn may be affected by their conceptualisation and choice of statistical methodology. It is helpful to consider whether hypotheses can be tested in more than one way and to establish if the conclusions remain broadly similar across different approaches used.

CONCLUSIONS

This chapter sought to provide a critical up-to-date review of the terms Mediation, Moderation, and Interaction as they are being commonly defined, discriminated, used, and tested as of 2013 and to consider some of their implications for quantitative educational research. That said, with a greater consideration on the historical origins of these terms as well as ‘real-life’ difficulties and ambiguities, we have also tried to equip educational researchers with the working knowledge to use these ideas with greater precision and clarity in their own research and to raise awareness of the broader substantive and methodological literatures which often vary in their chosen terminology.

From our review of both historic and current guidelines and practice, a number of recommendations emerge. First, it is essential that educational researchers have sufficient evidence to put forward and clearly distinguish one-tailed hypotheses such as Mediation and Moderation. Second, educational researchers must then gather (or have access to) data that is suitable to address these hypotheses with particular emphasis on the correct causal ordering of measures. Whether the quantitative research is correlational/survey or experimental in nature, for a one-tailed hypothesis to be adequately tested, there must be clear evidence of the appropriate temporal precedence between measures (as there was in the example provided in Section 4 above). Third, with the increasing availability, uptake, and sophistication of statistical software packages, it is our recommendation that quantitative researchers seriously consider the merits of Structural Equation Modelling (SEM) programmes such as EQS, LISREL, AMOS, and MPLUS. Not only have many of the historic statistical methods for testing Mediation, Moderation, and Statistical Interaction been incorporated into these packages, but they also facilitate the testing of these hypotheses when they are chained-together (e.g. multiple mediations as “indirect effects”) and combined (e.g. mediated-moderation) in ways that allow the researcher to address and model interesting and complex topics in educational contexts.

Finally and with the aims of this chapter aside, the critical reader might ask themselves, “*given the difficulties with these terms, are they really worth all the trouble?*” and this is an understandable question. Perhaps unsurprisingly, it is our opinion that the greater adoption and use of all of these terms by educational researchers is of direct benefit to the substantive knowledge of the field. In particular, the hypotheses of Mediation and Moderation are tools that give the educational researcher the ability to specify increasingly complex *and yet still testable* research hypotheses and enable them to explore causality in more plausible ways in complex and messy educational and social research contexts. Thus, Mediation and Moderation are tools that may empower the researcher to specify and test a greater number of clear hypotheses and, done with awareness and defensibly (if not the largely unobtainable “correctly”), this fosters the development of pyramid(s) of knowledge that can help to advance the knowledge base and possibilities of studying important educational research questions.

REFERENCES/BIBLIOGRAPHY

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*, 1173-1182.

Borkowski, J. G., Whitman, T. L., Wurtz-Passino, A., Rellinger, E. A., Sommer, K., Keogh, D., et al. (1992). Unraveling the “new morbidity”: Adolescent parenting and developmental delays. In N. Bray (Ed.), *International review of research in mental retardation* (Vol. 18, pp. 159–196). New York: Academic Press.

Burchinal, M., Peisner-Feinberg, E., Bryant, D., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science, 4*(3), 149-165.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modelling. *Journal of Abnormal Psychology, 112,* 558–577.

Creemers, B. P. M., Kyriakides, L., & Sammons, P. (2010). Background to Educational Effectiveness Research *Methodological Advances in Educational Effectiveness Research*. New York: Routledge.

Currie, J. (2001). Early Childhood Education Programs. *The Journal of Economic Perspectives, 15*(2), 213-238.

Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review, 85*, 341-364.

Edwards, J. R., & Lambert, L. S. (2007). Methods for Integrating Moderation and Mediation: A General Analytical Framework Using Moderated Path Analysis. *Psychological Methods, 12*(1), 1-22.

Elliot, C. D., NFER-NELSON, Smith, P., & McCulloch, K. (1996). *British Ability Scales Second Edition (BAS II)*. Windsor: NFER-NELSON.

Essex, M. J., Kraemer, H. C., Armstrong, J. M., Boyce, W. T., Goldsmith, H. H., Klein, M. H., et al. (2006). Exploring Risk Factors for the Emergence of Children’s Mental Health Problems. *Archives of General Psychiatry, 63*, 1246-1256.

Fergusson, D. M., & Lynskey, M. T. (1993). Maternal age and cognitive and behavioural outcomes in middle childhood. *Paediatric and Perinatal Epidemiology, 7*(1), 77-91.

Frazier, P. A., Barron, K. E., & Tix, A. P. (2004). Testing Moderator and Mediator Effects in Counselling Psychology. *Journal of Counselling Psychology, 51*(1), 115-134.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233-239.

Glass, N. (1999). Sure Start: the development of an early intervention programme for young children in the United Kingdom. *Children & Society, 13*(4), 257-264.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. London: Charles Griffin and Co.

Goldstein, H. (2003). *Multilevel Statistical Models*. Third Edition. London, Edward Arnold.

Goodnight, J. A., Bates, J. E., Staples, A. D., Petit, G. S., & Dodge, K. A. (2007). Temperamental Resistance to Control Increases the Association Between Sleep Problems and Externalizing Behavior Development. *Journal of Family Psychology, 21*(1), 39-48.

Hall, J., Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2009). The role of pre-school quality in promoting resilience in the development of young children. *Oxford Review of Education, 35*(3).

Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, Revised Edition*. New York: Teachers' College Press.

Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs, 76*(4), 408-420.

Hinshaw, S. P. (2002). Intervention research, theoretical mechanisms, and causal processes related to externalizing behavior patterns. *Development and Psychopathology, 14*, 798-818.

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping, and Proxy Risk Factors. *American Journal of Psychiatry, 158*, 848-856.

Luyten, H., & Sammons, P. (2010). Multilevel Modelling. In B. P. M. Creemers, L. Kyriakides & P. Sammons (Eds.), *Methodological Advances in Educational Effectiveness Research*. New York: Routledge.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the Practice of Dichotomization of Quantitative Variables. *Psychological Methods, 7*(1), 19-40.

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*, 83-104.

Masten, A. S. (2001). Ordinary Magic: Resilience Processes in Development. *American Psychologist, 56,* 227-238.

Masten, A. S. (2007). Resilience in developing systems: Progress and promise as the fourth wave rises. *Development and Psychopathology, 19*(3), 921-930.

McClelland, G. H., & Judd, C. M. (1993). Statistical Difficulties of Detecting Interactions and Moderator Effects. *Quantitative Methods in Psychology, 114*(2), 376-390.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When Moderation Is Mediated and Mediation Is Moderated. *Journal of Personality and Social Psychology, 89*(6), 852-863.

Muthén, L. K., & Muthén, B. O. (2010). Mplus software (Version 6). Los Angeles, CA: Authors.

NICHD. (2000). The interaction of child care and family risk in relation to child development at 24 and 36 months. *Applied Developmental Science, 6*(3), 144-156.

Nicholson, R. A., Hursey, K. G., & Nash, J. M. (2005). Moderators and Mediators of Behavioral Treatment for Headache. *Headache, 45*, 513-519.

Preacher, K. J., & Hayes, A. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*(4), 717-731.

Preacher, K. J., Rucker, D. D., Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42,* 185-227.

Reynolds, A. J., & Ou, S. (2003). Promoting Resilience through Early Childhood Intervention. In S. S. Luthar (Ed.), *Resilience and Vulnerability: Adaptation in the Context of Childhood Adversity*. Cambridge: Cambridge University Press.

Roe, R. A. (2012). What is wrong with mediators and moderators? *The European Health Psychologist, 14*(1), 4-10.

Rose, B., Holmbeck, G., Coakley, R., & Franks, E. (2004). Mediator and moderator effects in developmental and behavioral pediatric research. *Journal of Developmental and Behavioral Pediatrics, 25*, 1-10.

Rutter, M., & Silberg, J. (2002). Gene-Environment Interplay in Relation to Emotional and Behavioral Disturbance. *Annual Review of Psychology, 53*, 463-490.

Saunders, D. R. (1955). The 'moderator variable' as a useful tool in prediction. *Proceedings of the Conference on Testing Problems. Educational Testing Service*, 54-58.

Sobel, M. E. (1982). Aysmptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–212). San Francisco: Jossey-Bass.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2010). *Early Childhood Matters: Evidence from the Effective Pre-school and Primary Education project*. Abingdon: Routledge.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2012). *Effective Pre-school, Primary and Secondary Education 3-14 Project (EPPSE 3-14) - Final Report from the Key Stage 3 Phase: Influences on Students' Development form age 11-14*: Department for Education.

Talamini, R., Bosetti, C., La Vecchia, C., Dal Maso, L., Levi, F., Bidoli, E., et al. (2002). Combined effect of tobacco and alcohol on laryngeal cancer risk: a case-control study. *Cancer Causes and Control, 13*, 957-964.

Tatsuoka, M. M. (1973). Multivariate Analysis in Educational Research. *Review of Research in Education, 1*, 273-319.

Wu, A. D., & Zumbo, B. D. (2007). Understanding and Using Mediators and Moderators. *Social Indicators Research: An International Interdisciplinary Journal for Quality of Life Measurement.*