Observational modeling effects for movement dynamics and movement outcome measures across differing task constraints: A meta-analysis.

Derek Ashford
Institute for Biophysical and Clinical Research into Human Movement,
Manchester Metropolitan University, UK.

Simon Bennett
Department of Sports and Exercise Science, Liverpool John Moores University, UK.

Keith Davids
School of Physical Education,
University of Otago, New Zealand.

Key Words: Observational Modeling, Movement Dynamics, Movement Coordination, Movement Outcome, Task Constraints, Meta-analysis.
A meta-analysis was conducted on the observational modeling (OM) literature to quantify overall between-participant treatment effects obtained when acquiring movement behaviors. To evaluate predictions of the Visual Perception theoretical perspective on OM, effects were obtained and reported separately for movement dynamics (MD) and movement outcome (MO) measures. Overall mean OM treatment effect was $\delta_{bi}^u = 0.77$ for MD, and $\delta_{bi}^u = 0.17$ for MO measures. For both measures these effects reflected a significant advantage of OM over practice-only control conditions. Importantly, the magnitude of effects obtained was far stronger for MD compared to MO measures, confirming a distinctive response to OM during motor learning. The advantage for MD measures over OM measures was replicated for different types of task. OM was particularly beneficial for serial tasks ($\delta_{bi}^u = MD = 1.62$ and MO = 0.61). There were slightly reduced effects for continuous tasks ($\delta_{bi}^u = MD = 1.01$ and MO = 0.51), and smaller to medium sized effects for discrete tasks ($\delta_{bi}^u = MD = 0.56$ and MO = 0.10). These findings are in line with tenets of the Visual Perception perspective for observational modeling, which suggests demonstrations primarily convey relative motions required to approximate modeled movement behaviors.
Imitation of a model such as a parent, teacher or coach demonstrating a movement pattern is a common process in everyday life, intended to facilitate motor skill acquisition (De Maeght & Prinz, 2004). Most often called ‘observational modeling’ or ‘observational learning’, although sometimes referred to as vicarious learning, demonstration, social facilitation, mimicry, copying, and matched-dependent behavior (Williams, Davids, & Williams, 1999), this is a process whereby an individual assimilates through visual perception the information necessary to approximate the actions of others. From its early rise in popularity from social psychological research on imitation (e.g., Humphrey, 1921; Miller & Dollard, 1941; Ruffa, 1937; Bandura, 1962, 1969) to a more recent resurgence in areas such as neurobiology, robotics and artificial intelligence (e.g., Schaal, 1999; Schaal, Ijspeert, & Bilard, 2003), the issue of observational modeling has received significant empirical examination. Despite an abundance of work on observational modeling during the past 80 years, consistent findings have tended to be observed from studies that have used a small range of tasks (e.g., Bachman ladder climbing, arm movements), leading investigators to periodically highlight the need to explain equivocal results reported elsewhere within the literature (e.g., Gould, 1978; McCullagh & Little, 1989; Williams, 1984).

In response, there have been several qualitative reviews, which have highlighted a number of factors that could explain the equivocality, including the dependent measure used to reflect modeling effects (Wagman & Monroe, 1996), the task constraints present during motor learning (Gould & Roberts, 1982), and the age of the population being studied (Weiss, 1983). Although such reviews have been useful in stimulating discussion and further empirical study, they are susceptible to
potential subjective biases inherent in qualitative integration procedures, particularly
where inclusion criteria or the nature and extent of search procedures used are not
clearly stated. In addition, where qualitative reviews are not underpinned by a strong
theoretical rationale for predicting modeling effects, further subjective bias is possible
as the researcher is left to decipher and interpret the relevance of significant and/or
non-significant treatment effects. For a mature literature base such as that on
observational modeling, which contains varied empirical research designs, it is widely
accepted that an unbiased review can be obtained through a quantitative synthesis
(i.e., meta-analysis) of the reported (i.e., published and unpublished) effects. In
addition to quantifying the treatment effect across all studies, meta-analysis can also
be informed by theoretically-derived hypotheses regarding what factors might
influence the treatment effect. In this respect, meta-analysis can be a useful
supplement to qualitative reviews by providing a more objective test of a particular
theoretical account.

The Visual Perception perspective advocated by Scully and Newell (Scully,
1986; 1987; 1988; Scully & Newell, 1985) provides one such theoretically-based
account of the visual processes underlying observational modeling, which places
specific emphasis on the nature of perceptual information picked up by observers for
use in movement production. According to this view, changes in motor behavior
following observational modeling depend on the perception of “relative motion
information”, which is principally used by observers to guide their assembly of stable
patterns of coordination associated with a particular movement activity. The term
relative motion refers to the specific spatial-temporal relationships between and
within limbs, as well as the organization of the performer’s limbs relative to the
Observational Modeling

surrounding environment. It has been shown in the biological motion literature (e.g., Johansson, 1973, 1975), to be the principle source of visual information used by observers to identify and classify different types of human movement activities such as walking, cycling, and gymnastic moves.¹

The Visual Perception perspective on observational modeling is also clearly linked to Newell’s (1985) framework of motor learning stages. This proposes that during skill acquisition, early learning requires the search for and assembly of a functional coordination pattern, while later stages involve exploration of the control of key parameters of an established coordination pattern (Newell, 1985). The Visual Perception perspective on observational learning (Scully & Newell, 1985) advocates that the observer perceives information about the relative motion of a demonstrated action (e.g., the movement between hands in an object manipulation task), which then acts as informational (Warren, 1988, 1990) or instructional constraints (Newell & McDonald, 1992) on the learner’s organization of a functional coordination pattern. Moreover, it is suggested that visual demonstrations facilitate the early stage of skill learning because they convey relative motion information essential for the assembly of a novel or unfamiliar coordination pattern (Scully & Newell, 1985). In contrast, later in learning when the goal of the task is to flexibly adapt an existing, stable coordination pattern, learners are more focused on optimal scaling of the movement pattern relative to important environmental objects, surfaces and events (Cutting & Proffitt, 1982). Therefore, for more advanced learners, although relative motion can convey information associated with control variables (e.g., force and speed), modeling effects may be less pronounced. Instead, at the ‘control’ stage of learning there is an increased need for information from physical practice and knowledge of results for
ongoing refinements.

While the Visual Perception perspective makes clear predictions regarding the effectiveness of observational modeling at different stages of motor learning, its theoretical tenets raise an interesting question concerning whether there has been a bias in the extant literature on observation modeling to over-emphasize performance measures of movement outcomes (e.g., performance accuracy or error scores) rather than measures of movement dynamics that reflect movement (i.e., coordination) approximation (e.g., Al-Abood, Davids, & Bennett, 2001; Schoenfelder-Zohdi, 1992; Whiting, Bijlard, & den Brinkler, 1987). Such a distinction was noted in earlier research on modeling effects with a Bachman ladder task, where it was shown using qualitative techniques that movement dynamics (i.e., form) was a better indicator of modeling effects than movement outcomes (Feltz, 1982). This finding was later confirmed by McCullagh and Little’s (1989) investigation of the same task using quantitative methods. Their study showed that superior movement form approximation followed observational modeling, although no statistically significant differences between modeling and control conditions were found for movement outcome measures. The implication is that although observational modeling should be more effective during early stages of motor learning when participants are attempting to assemble an appropriate pattern of coordination, positive effects might be more likely to be evidenced by measures of movement dynamics (i.e., form) rather than those for movement outcomes.

To date, the distinction highlighted by the Visual Perception perspective between movement dynamics and movement outcomes in the observational modeling literature, and whether this might account for apparent equivocality in research
findings, has been the subject of conjecture and qualitative review. Therefore, the aim of the present study was to provide a quantitative synthesis of the effects of observational modeling on movement outcome and movement dynamic measures through the use of meta-analysis procedures.

Methods

Operational definitions

A randomized model was employed for the synthesis because it is a conservative approach which can cope with additional uncertainty arising from variations in experimental contexts, treatments, and procedures (see Cooper & Hedges, 1994). Studies under review were viewed as being representative of the larger population of investigations of movement-based observational modeling effects (which would include studies that reported such findings but otherwise failed to meet inclusion criteria within the current analysis). The analysis provided an integrative review to summarize between-participant effects across the observational modeling domain related to motor skill acquisition, consistent with Glass’s (1976) broad conception of an area review. Glass’s (1976) liberal inclusion criteria were adopted so that the contributions and effects contained within published and unpublished sources could be accounted for. Findings from these primary studies were converted into a common metric, $g$ (i.e., the standardized mean differences based on raw score dispersion data). Where multiple dependent measures based on the same construct (i.e. where the dependent measures are linearly correlated) were reported within a single primary source, only an average of these estimates was aggregated with effects from $k$ independent studies. Effect size estimates from each primary study were corrected for unreliability associated with small sample bias. Approximate weighted
data pooling and heterogeneity testing procedures as recommended by Hedges (1982) were employed to overcome concerns with the use of averaged effect size estimates, which may be derived from widely heterogeneous primary studies. Hedges’ (1983) heterogeneity test statistic has a chi-squared distribution with $k-1$ degrees of freedom, and was used to test the heterogeneity of pooled estimates.

Heterogeneous outcomes suggest that the variability of pooled estimates does not result solely from sampling variance but is likely to be influenced by at least one additional moderating variable (Hedges, Cooper & Bushman, 1992). Therefore, a within-participant (WP) $Q$-statistic was also calculated for individual effect size estimates within a pooled sample to provide an indication of the extent to which each estimate contributed toward the heterogeneity of a pooled sample (i.e., higher values indicate an increased contribution). This procedure permitted identification and subsequent removal of these outlying estimates. Removed estimates were analyzed and reported separately as sub-groups with a descriptive appraisal of each study’s primary characteristics to clarify potential features or design characteristic contributing to these outlying estimates (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Thomas & Nelson, 1996).

The literature search

A thorough review of the literature was conducted involving published and unpublished primary sources to avoid potential retrieval bias. Databases reviewed included: Medline, SPORTDiscus, PsycINFO, ScienceDirect, CatchWord, Ingenta, Ovid, and Dissertation Abstracts International. Search terms included: observational learning, observational modeling, demonstration(s), instruction, instructional method(s), modeling, vicarious learning, social facilitation, mimicry, matched-
dependent behavior, and copying. Searches were limited to human investigations over
the full period covered by the various reference sources. Because many early
observational investigations were directed towards social behavior modification
contexts (Bandura, 1962, 1969; Miller & Dollard, 1941; Piaget, 1951), the title often
suggested their subsequent exclusion. However, these primary studies were still
collated and reviewed carefully in order to scrutinize their reference pages for further
studies that could be appropriate for the current synthesis.

Eligibility criteria

The foremost criterion for inclusion was that the primary investigation
reported the effects of observational modeling in terms of movement dynamics (e.g.,
movement coordination, form approximation) and/or movement outcome (e.g.,
outcome goal). Movement dynamic measures included qualitative (e.g., judged
evaluations of movement form) and quantitative (e.g., kinematic data) measures
describing the degree of approximation towards the model’s movement dynamics.
Outcome measures included a range of dependent variables (e.g., accuracy, speed)
which were always directly associated with completion of the task goal. Additional
criteria for inclusion were: (i) studies had to provide sampling data (i.e., participant
numbers for all conditions); (ii) independent-groups post-test designs had to report
dispersion data (i.e., group means, standard deviations) or inferential statistics (e.g.,
independent t values); and (iii), single group (i.e., no control) and independent-groups
(i.e., observation and control conditions) repeated measures designs needed to report
dispersion data or appropriate inferential statistics (e.g., dependent t values) for the
conditions of interest. Wherever possible within repeated measures designs, raw score
data were used to determine rho values for the dependency within each group’s pre-
post treatment data, so that this statistic could be incorporated into each group’s effect size variance estimate. Although retention data are generally viewed as a better indicator of observational learning effects, their use across the literature is far less prevalent than data from the acquisition phase of learning studies following modeling treatments. Therefore, to maximize pooled estimates within the current analysis all estimates reflect the effects of observational modeling between baseline (e.g., pre-test) and acquisition test periods.

Data extracted from primary studies

Data pertaining to the following features were recorded: (i) authorship (e.g., names, institution) and primary study source (e.g., journal or book name, publication year, volume, issue, and page details); (ii) sampling methods (e.g., details regarding sample numbers, gender, age, and how samples were assigned and distributed across conditions). These details described the population from which the sample was drawn permitting consideration and recognition of any special population characteristics; (iii) experimental design (e.g., independent-groups post-test and single or independent-groups pre-post-test designs); (iv) dependent (e.g., typically movement outcome or movement dynamics measures) and independent (e.g., typically an observational modeling demonstration or a discovery control condition) variables; (v) demographic information about the model (e.g., gender, age, level of expertise) viewing methods and procedures (e.g., live or video model, viewing size, number and duration of presentations); and (vi) additional time factors common to all conditions (e.g., number of practice trials, the duration of the experiment, and the number and duration of practice and observation sessions).

Where possible, raw data for individual participants were coded to facilitate
implementation of meta-analytic procedures (e.g., calculation of rho values when not reported). If this was not possible, raw group data were coded. If dispersion data were not reported, statistical outcomes (e.g., F, or t statistics) were recorded to permit effect size estimation for the desired comparisons. The number of initial treatment estimates extracted from each independent study varied depending on the experimental design used and whether movement coordination and/or movement outcome dependent measures were reported. For example, an independent-groups design could produce two between-participant estimates (e.g., one movement outcome and one for movement dynamics). The number of between-participant estimates remained the same for independent-groups, repeated-measures designs. However, in these instances separate modeling and control within-participant estimates were used to determine each between-participant estimate.

Statistical procedures used for independent-group (post-test) designs.

Individual effect size estimates were calculated using Hedges and Olkin's (1985) pooled group standard deviation denominator. This initial biased estimate was corrected for small sample bias utilizing Hedges and Olkin’s (1985) correction procedure. Sample sizes commonly fell below 20 in many of the included observational modeling studies and sometimes below 10 where movement dynamics measures were reported. Therefore, an exact sampling variance equation was calculated (Morris & DeShon, 2002) rather than one based on a large sample approximation (Cooper & Hedges, 1994). To determine if pooled estimates derived from k independent studies were indeed representative of the same treatment effect, Hedges and Olkin's (1985) homogeneity test was conducted to determine whether differences in pooled effects were the result of sampling error alone or additional
experimental design characteristics. Where pooled effects were heterogeneous,
homogeneity values were determined for each independent effect size estimate to
determine outlying effects. Sub-groups of outlying effects were formed and
quantitative analysis continued. Where this was not possible, such effects were
qualitatively reviewed to determine what study characteristics might have contributed
to these extreme results. Where pooled effects from independent studies were
homogeneous, an overall weighted mean treatment effect was calculated (Hedges &
Olkin, 1985). The sampling variance for the weighted mean effect size estimates was
determined and used to calculate confidence intervals allowing interpretation of the
population effect.

Procedures for single and independent-groups repeated-measure designs.

Individual pre-post treatment effect size estimates for each dependent group
(observational modeling or control group) were determined utilizing the particular
group’s pre-test standard deviation as the denominator in the effect size equation
(Becker, 1988). The correction for small sample bias was conducted as for
independent-group designs, with $n - 1$ degrees of freedom (Hedges & Olkin, 1985).
An exact variance calculation was used (Morris, 2000) to determine the sampling
variance of each effect size estimate, which is the corrected representation of the
equation reported in Becker’s (1988) paper. This procedure incorporates a $\rho$ value
for the dependency or correlation between the pre- and post-test results for each effect
size estimate. B. J. Becker (personal communication, March, 8, 2002) confirmed that
the use of a substituted mean value of $\rho$ would offer a viable method to address the
dependency issue where $\rho$ values were unobtainable. Available $\rho$ values were
extracted from primary studies, collated and their distribution determined. Analysis
showed that these \( \rho \) values were normally distributed, Shapiro-Wilk (31) 0.149, \( p > 0.149 \). Therefore, a mean value of \( \rho \) was used to determine the sampling variance for individual effect sizes for studies where exact \( \rho \) was not available (Becker, 1988; Dunlap, Cortina, Vaslow & Burke, 1996; Morris, 2000; Morris & DeShon, 2002).

Finally, the repeated-measures (i.e., within-participant) effects for the experimental (i.e., modeling) and control conditions were used to determine the between-participant treatment effect (\( g_{\text{between}}^2 \)) using the procedure recommended by Becker (1988). This procedure subtracts the within-participant control group’s treatment estimate from the observational modeling group’s estimate. The variance for the between-participant treatment effect was determined by summing the variance values obtained for each within-participant estimate. On some occasions, single group, repeated-measure designs were reported that did not include a control condition. In such circumstances Becker (1988) recommended either utilizing a zero value to determine the overall independent-group’s treatment effect, or a mean effect size value obtained from studies that did include control conditions.\(^2\) The mean value used was based on a normally distributed sample of the reported effect size estimates for movement dynamics, Shapiro-Wilk (10) 0.935, \( p > 0.480 \), and movement outcome, Shapiro-Wilk (13) 0.945, \( p > 0.506 \). Therefore, a substituted mean of reported control effect size estimate values were used for studies that did not include a control condition. The aggregation of between-participant estimates from independent group designs, and between-participant estimates derived from within-participant reported pairs, with or without between-participant estimates from supplemented mean control-modeling pairs, provided the primary indication of overall observational
modeling treatment effects.

Results

Primary studies analyzed and those omitted

Of the 291 observational modeling studies initially considered, 227 were rejected because they failed to meet the eligibility criteria. The remaining 64 studies, involving 68 independent experiments were incorporated into the current meta-analysis producing 105 individual between-participant treatment estimates (MD = 33 and MO = 72).

Mean between-participant treatment effects for observational modeling

Movement dynamics measures. The chi square analysis indicated that the initial aggregation of MD estimates was heterogeneous (see Table 1). Sequential removal of five outliers resulted in a homogeneous sample $\chi^2(27, N = 732) = 37.40, p > 0.09$ and a strong overall mean treatment effect ($\delta_{\mu} = 0.77, \sigma^2_E(\delta) = 0.01, CI_{z/2} = \pm 0.24$). The frequency distribution of estimates contributing to this overall treatment effect is shown on the left-hand side of Figure 1 The pooled outliers were analyzed and found to be heterogeneous producing a far larger overall mean effect ($\delta_{\mu} = 1.85, \sigma^2_E(\delta) = 0.04, CI_{z/2} = \pm 0.53$) than that obtained for the homogeneous sample. This heterogeneity could not be attenuated through removal of further estimates. The presence and increased magnitude of these outlying estimates indicates the likely influence of at least one additional variable other then observational modeling. Table 2 provides sample sizes, source details, and selected characteristics associated with all movement dynamics effect estimates extracted from the literature including those removed as outliers.
Movement outcome measures. The chi square analysis indicated that the initial aggregation of movement outcome estimates was heterogeneous (see Table 1). Sequential removal of nine outlying estimates resulted in a homogeneous grouping $\chi^2(62, N = 2811) = 79.44, p > 0.07$ and a small overall mean treatment effect ($\delta_{li}^u = 0.17, \sigma^2_E(\delta) = 0.00, CI_{1.96} = \pm 0.11$). The frequency distribution for estimates contributing to this overall treatment effect is shown on the right-hand side of Figure 3. Analysis of the outlying estimates produced two homogeneous sub-groups. The initial subgroup required the removal of three estimates to obtain a chi square value of $\chi^2(5, N = 1259) = 6.01, p > 0.31$ and an overall mean effect of ($\delta_{li}^u = 1.13, \sigma^2_E(\delta) = 0.00, CI_{1.96} = \pm 0.16$). A further estimate was removed from the remaining pooled estimates to produce a chi square value of $\chi^2(1, N = 865) = 0.48, p > 0.49$ and an overall mean effect of ($\delta_{li}^u = 3.00, \sigma^2_E(\delta) = 0.01, CI_{1.96} = \pm 0.26$). Here again the considerable difference in magnitude of effects gained for the primary observational modeling sample compared to that for the much small outlying samples suggests these effects can not reliably be attributed to observational modeling + practice alone, but are likely to result from at least one additional influencing variable. Table 3 provides sample sizes, source details, and selected characteristics associated with all movement outcome effect sizes extracted from the literature including those removed as outliers.

Outlying estimates

As is the convention in meta-analyses, removed effects were scrutinized to determine possible reasons for their lack of alignment with other scores in the original sample. Four of the five movement dynamics outliers represented disproportionately large effect estimates (i.e. Dubanaski & Parton, 1971 [two estimates]; Lugana, 1996;
Observational Modeling

Magill & Schoenfelder-Zohdi., 1996), and one a negative estimate (Wiese, 1989), respectively. Dubanaski and Parton’s (1971) object manipulation investigation produced the largest movement dynamics outliers \( g_{\text{bi}}^u = 13.27 \) and \( 5.20 \), where \( g_{\text{bi}}^u \) describes a unbiased between-participant treatment estimate derived from a single independent investigation. The effects for reproducing a sequence of temporal and spatial arm movements (Laguna, 1996) and a rhythmic dance sequence (Magill, & Schoenfelder-Zohdi, 1996) were, in comparison, less extreme \( g_{\text{bi}}^u = 3.52 \) and \( 2.02 \) respectively). In contrast, a modified softball-pitching task produced a negative outlier \( g_{\text{bi}}^u = – 0.01 \) in Wiese’s (1989) investigation. A number of factors that may have contributed to these outlying effects were considered, although no single variable appeared to provide a satisfactory explanation given the limited outlier sample. For example, increases in effect size estimates from these investigations did not simply result from performing more practice trials; large and small effects were obtained from studies using relatively few practice trials. Still, it is interesting to note that the only negative effect resulted from a discrete task, whilst the remaining highly positive outliers were obtained for serial tasks.

A qualitative review of nine disproportionately large movement outcome (i.e., goal) outliers revealed that four were derived from studies that used atypically large sample sizes \( (n = 135 \) and \( 45, \) Landers, 1975; \( n = 424 \) and \( 417, \) Roshal, 1949 [three estimates]). However, large effects (including the other 5 outliers) were also obtained from studies using more typical sample sizes, hence indicating that sample size would not appear to account for the increased effect estimates. Landers (1975) and Landers and Landers (1973) investigations produced two outliers \( g_{\text{bi}}^u = 1.33 \) and \( g_{\text{bi}}^u = 0.93, \)
respectively) using a Bachman ladder climbing task. As can be seen in Tables 2 and 3, this task has been particularly popular, and has produced consistent between-participant treatment effects (Feltz, 1982; Feltz & Landers, 1977; Lirgg & Feltz, 1982). It is not obvious, therefore, why such large effects were obtained from these two studies.

Outliers were also obtained for three-knot tying tasks (Bowline $g_{bi}^u = 1.09$; Sheetbend $g_{bi}^u = 3.80$ & Spanish Bowline $g_{bi}^u = 2.99$) reported by Roshal (1949; 1961). Such tasks are atypical within the observational modeling literature, and involve considerable fine movement coordination and control to produce a complicated and specific sequence of movements. For this reason they have also been recognized as being particularly difficult to verbalize and describe (Annett 1986), possibly leading to increased benefits from observational modeling. Finally, in common with movement dynamic effects, only one discrete task (golf swing) was represented within the outlying movement outcome estimates (Nelson, 1958).

This initial qualitative review indicated that outliers were typically represented by disproportionately large effect estimates. Furthermore, although it is difficult to attribute differences in effect size magnitude to a single factor based on the scrutiny of a relatively small number of outliers, there was some initial evidence that abnormally large positive effects were often obtained from studies using serial tasks. This is not the first time it has been recognized that different task constraints might influence the effects of observational modeling procedures (Al-Abood, 2001; Gould, 1978; McCullagh et al., 1989; Williams et al., 1999; Williams, 1984). However, such conclusions have often previously been based on a relatively small number of empirical studies that compared effects obtained from specific tasks (e.g., Gould,
Observational Modeling

1978; Martens, Burwitz & Zuckerman, 1976). Having reviewed the extant literature on observational modeling for the purposes of the initial meta-analysis, there is the additional opportunity to produce an unbiased quantification of observational modeling effects obtained for different types of tasks. In doing so, it is important to reliably classify tasks according to the different demands they place on the individual. While this is by no means a straightforward undertaking (see Gould, 1978), the preliminary observation of the outliers indicated that there could be some merit in pooling of estimates using the traditional and well accepted discrete, serial and continuous skill classifications (Schmidt & Lee, 1999; Magill, 1993).

Methods

Tasks used in primary studies included within the first analysis were classified by three motor behavior investigators as being discrete, serial or continuous. Task effects were determined for movement dynamic and movement outcome estimates independently. This produced six pooled estimate samples, which were analyzed to determine homogeneity and overall mean treatment effects (see first analysis methods). Tasks were only incorporated within pooled samples where complete agreement was obtained from all three classifiers.

Results

Primary studies analyzed.

Of the 105 between-participant treatment estimates obtained for the first analysis, 98 estimates were classified as involving discrete, serial or continuous tasks. Complete agreement was not obtained for seven movement outcome estimates and these were not included in the task analysis (see Table 3 estimates 66-72). An inspection of estimates within Tables 2 and 3 shows that discrete tasks have been used
more extensively in the observational modeling literature (MD = 18 and MO = 40),
whilst serial (MD = 7 and MO = 13) and continuous (MD = 8 and MO = 12) tasks
have been used less frequently. Additionally, only a limited range of serial and
continuous tasks have been utilized to examine the effects of observational modeling.
For example, approximately 80% of the estimates obtained for continuous tasks have
utilized Bachman ladder, slalom ski simulator or stabilometer tasks. With serial tasks,
55% of estimates reflected barrier knock down or multi-step motor sequence tasks.

Effects obtained for movement dynamic measures.

Table 2 provides sample sizes, source details, and selected characteristics
associated with individual MD estimates extracted from the literature including
outliers. The top, middle and bottom sections within Table 2 reflect discrete
(estimates 1-18), serial (estimates 19-25) and continuous (estimates 26-33) tasks
respectively.

Discrete tasks. The chi square analysis of pooled discrete task estimates
indicated a homogeneous sample $\chi^2(17, N = 498) = 26.98, p > 0.06$ and a medium
sized overall mean treatment effect ($\delta_{\text{BI}} = 0.56, \sigma^2_{E(\delta)} = 0.01, \text{CI}_{\alpha/2} = \pm 0.29$).

Serial tasks. The initial chi square analysis of pooled serial task estimates was
heterogeneous (see Table 1). Sequential removal of four outlying estimates resulted in
a homogeneous grouping $\chi^2(2, N = 102) = 4.49, p > 0.11$ and a slightly reduced
overall mean treatment effect ($\delta_{\text{BI}} = 1.62, \sigma^2_{E(\delta)} = 0.07, \text{CI}_{\alpha/2} = \pm 0.67$) compared to
the original heterogeneous sample. A sub-group analyses of the removed outlier
remained heterogeneous resulting in an extremely large overall mean effect ($\delta_{\text{BI}} =$
2.96, $\sigma^2_{E(\delta)} = 0.10, \text{CI}_{\alpha/2} = \pm 0.81$) compared to that obtained with the larger
Continuous tasks. The chi square analysis of pooled continuous task estimates indicated a homogeneous sample $\chi^2(7, N = 208) = 5.39, p > 0.61$ and produced a very large overall mean treatment effect ($\delta_{\text{Bi}} = 1.01$, $\sigma^2_{E(\delta)} = 0.03$, CI$\alpha/2 = \pm 0.43$).

Effects obtained for movement outcome measures.

Table 3 provides sample sizes, source details, and selected characteristics associated with all movement outcome effect sizes extracted from the literature including those removed as outliers (the top middle and bottom blocks within Table 3 reflect discrete (estimates 1-40), serial (estimates 41-53), continuous (estimates 54-65) and unclassified (estimates 66-72) tasks respectively.

Discrete tasks. The initial chi square analysis of pooled discrete task estimates was heterogeneous (see Table 1). However, homogeneity was achieved $\chi^2(38, N = 1740) = 36.61, p > 0.53$, following removal of one outlying estimate. This resulted in a slightly reduced overall mean effect ($\delta_{\text{Bi}} = 0.10$, $\sigma^2_{E(\delta)} = 0.00$, CI$\alpha/2 = \pm 0.10$) compared to the initial heterogeneous sample.

Serial tasks. The initial chi square analysis of pooled serial task estimates was heterogeneous (see Table 1). Removal of two outlying estimates resulted in a homogeneous grouping $\chi^2(10, N = 424) = 13.30, p > 0.21$, and a slightly increased overall mean effect ($\delta_{\text{Bi}} = 0.61$, $\sigma^2_{E(\delta)} = 0.01$, CI$\alpha/2 = \pm 0.27$) compared to the initial heterogeneous sample.

Continuous tasks. The initial chi square analysis of pooled continuous estimates was heterogeneous (see Table 1). Two outlying estimates were removed to obtain a homogeneous grouping $\chi^2(9, N = 619) = 14.04, p > 0.12$, and a slight
decreased overall mean effect ($\delta_{Bi} = 0.51$, $\sigma^2 E(\delta) = 0.01$, $CI_{\alpha/2} = \pm 0.27$) compared to the initial heterogeneous sample.

General Discussion.

Motivated by the predictions of the Visual Perception perspective, this quantitative review sought to examine whether equivocality in the reported effects following observational modeling has been introduced by an over-emphasis on dependent measures of movement outcomes over movement dynamics. To achieve this aim, a meta-analysis was performed on studies that met specified inclusion criteria to quantify the magnitude of treatment effects following observational modeling, over and above those gained through practice-only control conditions (i.e., the between-participant treatment effect). The overall mean treatment effect associated with observational modeling reported for movement dynamics data was $\delta_{Bi} = 0.77$ standard deviations greater than that obtained from the no-observation control conditions. This finding represented a significant advantage ($p < 0.01$) for modeling compared to physical practice-only conditions in approximating movement dynamics. In contrast, the overall mean treatment effect for movement outcome (i.e., performance outcome) measures was much more modest ($\delta_{Bi} = 0.17$), although the advantage for modeling over physical practice-only control was still statistically significant ($p < 0.01$). These findings show that while there are additional benefits of observational modeling over practice-only control conditions, the magnitude of the treatment effect is dependent on the measure used to quantify acquisition.

It is noteworthy, however, that because the overall treatment effect of observational modeling was significant for both movement dynamic and movement outcome measures, there are likely to be additional moderating variables that have
Observational Modeling contributed to the reported equivocality (i.e., negative or non-significant effects) in the extant literature. On the basis of a qualitative review of the outlying effects from the initial meta-analysis, a further meta-analysis was performed to quantify the additional influence of different constraints associated with serial, continuous and discrete task classifications. This additional analysis revealed that the positive treatment effect of observational modeling for movement dynamic measures was replicated across each of the different task constraints. Observational modeling produced a large treatment effect for serial tasks ($\delta_{Bi}^{u} = 1.62$), a slightly reduced effect for continuous tasks ($\delta_{Bi}^{u} = MD = 1.01$), and a medium sized effect for discrete tasks ($\delta_{Bi}^{u} = MD = 0.56$). A similar trend was found for movement outcome measures ($\delta_{Bi}^{u} = 0.61$ [serial], 0.51 [continuous], and 0.10 [discrete]), although the treatment effect for discrete tasks was particularly small. This more objective analysis revealed that the magnitude of the treatment effect for observational modeling is influenced by the dependent measure used to describe treatment effects, as well as the task constraints of the modeled movement. Studies reporting movement outcome measures derived from discrete tasks tended to report small or even negative treatment effects following observational modeling. They were more likely to produce contradictory equivocal? findings compared to the range of positive treatment effects reported in studies using serial or continuous tasks and quantifying acquisition or learning with movement dynamic or movement outcome measures. Interestingly, such studies represent a disproportionately large amount of the work on observational learning, and have therefore contributed to the confusion in understanding within this area.

In addition to providing an explanation for some of the equivocal findings in the extant literature on observational modeling, the current meta-analyses offer
indirect support for the predictions of Scully and Newell’s (1985) Visual Perception perspective. According to this approach, observation of a model provides relative motion information, which conveys the topographical characteristics associated with the modeled movement behavior. It is suggested that the perception of relative motion information enables individuals to approximate an observed movement, leading to learned changes in motor behavior. The implication, which was confirmed by the analyses reported in the current study, is that observational modeling will be particularly beneficial in guiding observers to assemble the pattern of coordination demanded by a particular movement activity. The Visual Perception perspective also predicts that observational modeling will be particularly facilitating during the early stage of skill learning, when it is important that visual demonstrations convey relative motion information that is unfamiliar to the novice learner. While the current study did not compare different stages of skill learning (sample selection for most studies is designed to highlight acquisition and learning effects), the finding that task constraints influenced the effectiveness of observational modeling might be considered in a similar light. Serial tasks require the coordination of multiple movements in a specific sequence (Laguna, 1996; $g^w_{bi} = 3.52$), and are likely to be more complex for the novice learner than discrete tasks that require a single case of coordinated movement (Weeks, 1992; $g^w_{bi} = -0.009$), or a continuous task requiring multiple cases of the same coordinated movement. By providing information about the required coordination for each individual movement (e.g., intra-limb coordination), as well the correct coordination between individual movements, observational modeling reduces the unfamiliarity of serial tasks that would be difficult to achieve through verbal
instructions (Annett, 1986), and thus enables large improvements in motor performance. It is interesting to note that although observational modeling led to improvements in the ability to coordinate movements under serial, continuous and discrete task constraints, this observation was not reflected by a similar magnitude of improvement in movement outcome. In fact, for discrete tasks, it was found that practice alone is equally as likely to lead to improvements in movement outcome measures as observational modeling and practice. Observation of Tables 2 and 3 indicates that the discrepancy between the effects of observational modeling for different task constraints might be explained by the dependency of movement outcome measures on movement dynamics. For example, with serial and continuous tasks, improvements in movement outcome (e.g., percentage of correct trials in 6-part motor sequence) might be most effectively achieved by adopting a particular pattern of movement coordination. However, with discrete tasks (e.g., basketball free-throw) there are potentially many more patterns of coordination that could produce a successful movement outcome (e.g., throw the basketball though the hoop). Consequently, even if the participant assembles a pattern of coordination similar to that of the visual demonstration, it does not follow that this strategy will result in an improvement in movement outcome. The implication, therefore, is that positive effects of observational modeling for measures of movement outcome, are most likely to occur when a specific pattern of movement coordination is required, which is available to be perceived from the relative motion information in the visual demonstration. A question for future research is whether positive effects of observational modeling for measures of movement outcome will be shown if a
discrete task imposes a ceiling effect that can only be overcome by assembling a
pattern of coordination similar to that of the visual demonstration. Additionally, while
susceptibility of serial tasks to OM effects was noted earlier in this paper, empirical
work is needed to examine a wider range of serial tasks to clarify whether reported
effects were biased by propensity of researchers to select only a limited range of
constraints in this category (e.g., barrier knock down tasks).

Summary

Rosenthal (1991) suggested that “meta-analytic reviews are more likely to lead
to summary statements of greater thoroughness, greater precision, and greater inter-
subjectivity or objectivity” (p.11). The analyses presented in this paper represent the
first quantitative reviews of the observational modeling literature associated with
motor learning. It was found that observational modeling is particularly effective for
the acquisition of movement dynamics, but more modest in attaining movement
outcomes, particularly when learning discrete tasks. This objective synthesis of the
observational modeling literature provides a robust methodological framework for
future investigations to design and interpret studies of modeling procedures.
References

*References prefixed by an asterisk indicate studies included in the meta-analysis.


Annette, J. (1986). On knowing how to do things. In *Generation and*


Observational Modeling


Morris, S. B. (2000). Distribution of the standardised mean change effect size for meta-analysis on repeated measures. *British Journal of Mathematics and


action in sport. (pp.238-273), E & FN Spon.


Observational Modeling

1. Bibliography of excluded observational modeling studies.


Cooper, L.K., & Rothstein, A.L. (1981). Videotaped replay and the learning of


Gallagher, J.D., & Thomas J.R. (1984). Rehearsal strategy effects on


consecutive demonstration and practice in the learning of a serial mechanical-
assembly task. In A.A. Lumsdaine (Ed.), *Student responses in programmed
demonstration and overt practice as a function of task organisation. In A.A.
Lumsdaine (Ed.), *Student responses in programmed instruction*. National Academy of
May, M.A. (1946). The psychology of learning from demonstration films.
*Journal of Educational Psychology*, 37(1), 1-12.
Journal of Sport Psychology*, 9, 249-260.
psychological considerations. In R. N. Singer and M. Murphey and L. K. Tennant
(Eds.), *Handbook of Research on Sport Psychology* (pp.106-126).
influence of model type on the learning of a free-weight squat lift. *Research Quarterly
for Exercise and Sport*, 68(1), 56-61.


College of Human Development and Performance, University of Oregon, Eugene, Ore, 2 microfiches (170 fr.): negative, ill.; 11 x 15 cm.


Sidaway, B., & Hand, M.J. (1993). Frequency of modeling effects on the


1 of Psychology, 43(2), 165-187.
5 Schorndorf.


Footnote

1. Whilst the Visual Perception perspective offers an acceptable explanation for the findings from the meta-analysis, as is often the case with empirical outcomes alternative theoretical approaches may be applied to interpret and explain these results.

2. An additional analysis was run to quantify the effect that would be obtained if Becker’s (1988) substitution procedures were not applied. Treatment effects for movement dynamics (MD) and movement outcome (MO) measures were comparable if slightly less conservative compared to those obtained using Becker’s procedures [e.g., Homogeneous MD $\chi^2(17, N = 456) = 24.51, p > 0.1062$, and MO $\chi^2(34, N = 1723) = 48.05, p > 0.0557$, pooled estimates produced overall mean treatment effects of $(\delta_{\mu} = 0.92, \sigma^2_E(\delta) = 0.01, CI_{95\%} = \pm 0.30)$ and $(\delta_{\mu} = 0.42, \sigma^2_E(\delta) = 0.06, CI_{95\%} = \pm 0.15)$ respectively].

3. The parameter gained from the meta-analysis remains a sample estimate, an approximation (i.e., the overall weighted mean effect size). Confidence intervals are used to illustrate the extent to which a sample estimate may be generalized to the population effect (i.e. the true effect size). Confidence intervals (95 or 99 %) are calculated around the obtained weighted mean effect size estimate to determine whether it encompasses zero, where confidence limits do not overlap zero, the result is statistically significant. Or put another way “the true value is unlikely to be zero, or that there is a real effect” (Hopkins, 2002).

4. Early suggestions that modeling fulfilled an informational function (Bandura, 1969; Sheffield, 1961) prompted Gould (1978) to investigate the effects of information load.
Gould defined informational load as, “the number of procedural steps an individual must execute to correctly perform the task, and the degree to which one particular performance strategy (one specific alternative set of behaviors which is not readily apparent) leads to correct task execution.” (p.24). Results were obtained for tasks classified as involving low (a ball snatch task), moderate (a rebound-ball-roll task), or high (a 7-piece geometric construction task) informational loads. Results indicated that task characteristics influenced modeling, with facilitative effects occurring with increased informational load and reduced movement novelty.

5. Abbreviation indicates use of non-dominant hand.