Title:
Identification of Causal Effects Using Gain Scores

Authors and Affiliations:
Yongnam Kim, University of Wisconsin-Madison
Peter M. Steiner, University of Wisconsin-Madison
Background
The use of gain scores for estimating the effect of an intervention has been frequently avoided in education and psychology. Campbell and Erlebacher (1970, p. 197) claimed, “gain scores are in general [...] a treacherous quicksand” and Cronbach and Ferby (1970, p. 80) even recommended researchers to “frame their questions in other ways.” Though a few more recent studies argue that gain scores can be effective for identifying causal effects in observational studies (e.g., Allison, 1990; Maris, 1998), there is still a widespread aversion to gain scores (Thomas & Zumbo, 2012).

Recent advances in causal inference, in particular the graphical models approach (Pearl, 2009; Spirtes, Glymour & Scheines, 2000), shed new light on the misunderstandings about gain scores. Graphical representations directly visualize the assumptions required for identifying a causal effect but also help in explaining the bias-removing mechanics of an identification strategy. Though the graphical approach is rather new to the social sciences, especially in education and psychology, there is a growing interest among researchers in applying graphical models to quasi-experimental designs or missing data problems (e.g., Steiner et al., 2015; Thoemmes & Mohan, 2015).

Focus of Study
This paper introduces graphical models for gain scores and develop graph-based arguments for the causal identification of the average treatment effect on the treated. The graphs allow practitioners to directly see the unique assumptions and bias-removing mechanics. More generally, this paper investigates the distinct role of pretests in different causal identification strategies.

Findings
Figure 1A shows the graphical model for a basic gain score approach where Z is the treatment like participating in a math camp, P the pretest, Y the posttest, and A an unobserved confounder like students’ ability. Given the graph, it is obvious that the causal effect of the math program on the posttest is not identified because the back-door path Z ← A → Y cannot be blocked (because A is unmeasured). Though one may try to match subjects on the pretest P or to control for P via covariance adjustment, such a conditioning cannot succeed in removing all the confounding bias.

However, gain scores can identify the causal effect under certain conditions. Using gain scores first requires the computation of the gain score G (G = Y – P) as shown in Figure 1A. Then the graph shows that the causal effect of Z on G is confounded by two non-causal paths: i) Z ← A → P → G, and ii) Z ← A → Y → G. If the confounder A affects the pretest P and the posttest Y to the same extent, that is, the impacts A → P and A → Y are the same, then the spurious associations via the two paths perfectly offset each other. This assumption is frequently called the common trend assumption (Lechner, 2010). Since the non-causal associations offset each other the average causal effect of Z on G is no longer confounded (but the effect of Z on Y is still confounded). The graph-based arguments can be used in the causal identification in more complex situations. For instance, when the pretest P affects either the treatment or outcome (Figures 1B and 1C; identification fails here) or the common measurement error affects the pretest and posttest simultaneously (Figure 1D; identification succeeds here because the non-causal paths are the same as in Figure 1A).
The graphs above also reveal the difference in the bias-removing mechanics of gain scores and conditioning methods (e.g., covariance adjustments or matching). While conditioning methods partially block the confounding path $Z \leftarrow A \rightarrow Y$, gain scores remove the confounding bias by offsetting the two non-causal paths. Recently, researchers have found that conditioning on a covariate can increase a bias due to bias amplification (Pearl, 2010; Steiner & Kim, 2016). However, because gain score methods do not condition on the pretest, bias amplification is not an issue. Figure 2 demonstrates that the bias in the treatment effect resulting from gain scores is not affected by bias amplification while conditioning methods can be seriously affected by it.

A drawback of gain scores is that the common trend assumption may not hold in practice, that is, the unobserved confounder’s ($A$) impact on the pretest $P$ and posttest $Y$ might differ. Nonetheless, adding additional design elements or other strategies may still allow us to identify the causal effect. In Figure 3A, two pretests $P_0$ and $P$ are available. Then, if $A$’s impact on the multiple measures only changes linearly over time, the differences in two gain scores (Figure 3B) results in a new version of the common trend assumption (Figure 3C). Second, though the simple gain score method does not require conditioning on any other covariate, conditioning on some covariates may make the common trend assumption more plausible (Figures 4A and 4B). Third, a linear or nonlinear transformation of the pretest and posttest can help to relax the assumption. Though it does not hold on the original scale, the assumption may hold after transforming the measures (Figure 5).

**Conclusion**

This paper visualizes the causal identification strategy of gain scores and highlights its distinct use of the pretest to deal with confounding bias. As the graphs show, gain scores remove the bias by offsetting non-causal associations and are not prone to bias amplification. The offsetting mechanics are based on the common trend assumption rather than the unconfoundedness assumption. The common trend assumption does not require a highly reliable pretest while reliable measures are required for the unconfoundedness assumption of conditioning methods. However, the common trend assumption is sensitive to the scaling of the pretest and posttest. The findings show that a “good” pretest for the gain score methods differs from a “good” pretest for matching or covariance adjustments. Finally, the gain scores’ identification strategy directly applies to a more general class of designs that includes comparative interrupted time series designs, difference-in-differences and fixed effects methods.
References
Appendix: Figures
Figure 1. Causal graphs for gain scores holding the common trend assumption (indicated by tick marks on the arrows). (A) Simple gain score graph. (B) Graph when pretest determines treatment. (C) Graph when pretest determines outcome. (D) Graph when a common measurement error affects pretest and posttest together.
Figure 2. Bias of gain score and conditioning estimators (e.g., matching or covariance adjustments), based on the data-generating model depicted in Figure 1A.
Figure 3. Causal graphs when multiple pretests are available. (A) Data generating model. (B) Computing two gain scores separately. (C) Simplified graph with respect to the two gain scores. (D) Fully descriptive graphs with multiple gain scores.
Figure 4. (A) Causal graph expressed with the merged confounder set $U$. (B) Identical causal graph but separating two components in $U$ into $A$ and $S$.

Figure 5. Causal graph with the transformed pretest $P^*$ and posttest $Y^*$. 