The Use and Impact of Adjusted R^2 Effects in Published Regression Research

Lesley F. Leach **Robin K. Henson**
University of North Texas

This paper empirically evaluates the reporting of adjusted effect sizes (e.g., adjusted R^2, omega^2) in published multiple regression studies by (a) documenting the frequency of adjusted effect reporting and interpretation, (b) identifying the types of corrected effects reported, and (c) estimating the degree of "shrinkage" present across regression analyses by using the information found in published journal articles to calculate corrected effects based on various formulae. Adjusted effects were infrequently reported in the literature, and interpretation of adjusted effects that were reported was rare.

Researchers are becoming increasingly aware that interpretation of effect sizes is critical in evaluating empirical results (Henson & Smith, 2000; Henson, 2006; Kirk, 1996; Rosnow & Rosenthal, 1989; Thompson, 1996; Thompson & Snyder, 1997). The American Psychological Association (APA) Task Force on Statistical Inference (Wilkinson & APA Task Force on Statistical Inference, 1999) stated:

> It is hard to imagine a situation in which a dichotomous accept-reject decision is better than reporting an actual *p*-value or, better still, a confidence interval. . . *Always* provide some effect-size estimate when reporting a *p*-value. (p. 599, italics added).

The Task Force went on to state, “Always present effect sizes for primary outcomes . . .It helps to add brief comments that place these effect sizes in a practical and theoretical context" (Wilkinson & APA Task Force on Statistical Inference, 1999, p. 599).

This directive was a substantial step beyond the fourth edition of the APA’s Publication Manual, which only *recommended* reporting of effect sizes in research (APA, 1994). Several empirical studies demonstrated, however, that this recommendation had little impact on the number of effect sizes reported in articles and it affected the interpretation of effect sizes even less (cf. Henson & Smith, 2000; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000).

The fifth edition of the APA manual (APA, 2001) incorporated the Task Force’s directive, stating “For the reader to fully understand the importance of your findings, it is almost always necessary to include some index of effect size or strength of relationship in your Results section” (p. 25). The current APA manual also called the "failure to report effect sizes" a "defect in the design and reporting of research" (p. 5). At least 23 journals have followed suit, requiring the inclusion of effect sizes with statistical results (Onwuegbuzie, Levin, & Leech, 2003).

The use of effect sizes has been widely discussed in the literature vis-à-vis null hypothesis significance tests (NHST). A discussion of issues surrounding the use of NHSTs is beyond the scope of this paper. Harlow, Mulaik, and Steiger (1997) present a balanced discussion of the debate for interested readers. Huberty and Pike (1999) and Huberty (2002) document the historical development of statistical testing and effect sizes, respectively.

Indeed, Pedhazur and Schmelkin (1991) noted that, “Probably few methodological issues have generated as much controversy among sociobehavioral scientists as the use of [statistical significance] tests” (p. 198). Elsewhere, Pedhazur (1997) indicated that the “controversy is due, in part, to various misconceptions of the role and meaning of such [statistical significance] tests in the context of scientific inquiry” (p. 26). These “misconceptions” have been attacked for considerable time (see e.g., Berkson, 1942; Tyler, 1931), and yet they persist in modern research practice (Cohen, 1994; Finch, Cumming, & Thomason, 2001). Nevertheless, current methodological practice is increasingly emphasizing the need for effect size indices and more accurate interpretation of NHSTs (Kline, 2004).

Some researchers recommend using effect sizes and NHSTs together (Fan, 2001; Huberty, 1987). Moreover, some critics of NHSTs have argued that effect sizes should be reported whether or not the results are statistically significant (Rosnow & Rosenthal, 1989; Thompson, 1999). As Roberts and Henson (2002) stated, “. . .one remaining point of debate concerns whether effect sizes should be reported (a) for all null hypothesis tests, even non-statistically significant ones, or (b) only after a finding is first determined to be statistically significant” (pp. 242-243).

Types of Effect Sizes: Corrected and Uncorrected Indices

There are many different effect size indices from which researchers can choose, but most can be grouped into two broad categories: (a) measures of standardized differences (e.g., Cohen's *d*, Hedges' *g*) and (b) variance–accounted-for measures (e.g., R^2, η^2) (Kirk, 1996; Kline, 2004; Olejnik & Algina, 2000; Onwuegbuzie, Levin, & Leech, 2003). Reviews of various effect size indices are provided by Olejnik and Algina (2000), Snyder and Lawson (1993), and Yin and Fan (2001).

In addition, effect sizes can be further classified as "uncorrected" or "corrected" measures (Thompson, 2002). For example, R^2 is commonly used in multiple regression applications and is the most prevalent effect size index documented in the literature – most likely due to the fact that practically all statistical computer packages routinely provide R^2 as part of regression output (Kirk, 1996). Studies have shown, however, that R^2 systematically overestimates the proportion of explained variance to total variance expected in the population or future samples (Carter, 1979; Fan, 2001; Snyder & Lawson, 1993; Thompson, 1990, 1999; Yin & Fan, 2001). That is, general linear model analyses such as multiple regression commonly utilize the ordinary least squares (OLS) estimation method to obtain the greatest possible effect size. Analyses using this estimation method capitalize on *all* the variance in a sample, including the variance attributable to sampling error that is unlikely to be present in future samples or the population (Thompson & Kieffer, 2000). Because the effect size accounts for error unique to the sample data, the resulting "uncorrected" R^2 is often found to be a biased estimate of the variance explained in the population (Roberts & Henson, 2002; Yin & Fan, 2001) and future samples (Thompson, 1990).

To statistically remove the bias associated with sampling error, various adjustment formulae can be used to "shrink" the effect size by the theoretical amount of sampling error present in a given sample (Snyder & Lawson, 1993). The amount of shrinkage is determined using the factors that affect sampling error. Theoretically, sampling error increases (a) as sample size decreases, (b) as the number of variables in the model increases (and, by extension the number of predictors increase), and (c) as the population effect decreases (Thompson, 1999; Vacha-Haase & Thompson, 2004). Because adjustment formulae limit the influence of the factors that increase sampling error, these "corrected" effects provide a better estimate of the population squared multiple correlation coefficient (Carter, 1979; Larson, 1931; Pedhazur, 1997). But as this paper demonstrates, corrected effects are rarely reported, and the failure to report such corrected effects may impact result interpretation.

Purpose

Because corrected effects can be more accurate estimates of the effect in the population or future samples, the purposes of the present study were to (a) document the frequency of corrected effect reporting and interpretation, (b) identify the types of corrected effects that are reported, and (c) estimate the degree of shrinkage present when authors do not give corrected effects. Information found in the reviewed articles was used to calculate corrected effects based on various formulae (Snyder & Lawson, 1993; Yin & Fan, 2001). This analysis facilitated inspection of interpretation differences resulting from effect size adjustment and permitted empirical investigation of the amount of correction provided by the various corrected effect formulae. Because R^2 and adjusted R^2 are typically reported with regression results in statistical software packages, this paper addressed only multiple regression applications.

Adjusted R^2 Formulae

There are many formulae available for calculating corrected effect sizes. Table 1 outlines various formulae presented by Snyder and Lawson (1993) and Yin and Fan (2001), which shrink R^2 based on the number of predictors (k), sample size (n), and the obtained effect (R^2) as an initial estimate of the population effect. The adjustment formulae fall into two different categories based on their purposes: (a) population effect estimates and (b) future sample effect estimates. Population effect estimates approximate the association strength expected to be realized in the population (Yin & Fan, 2001), while those in the future sample category estimate the effect likely to be found upon replication of the study with a new sample (Snyder & Lawson, 1993). One could expect greater shrinkage to be more likely with future sample estimates because "[they] must adjust for sampling error present in both the present study and some future study" (Snyder & Lawson, 1993, p. 340). Conversely, adjusted effect estimates of the population parameter only adjust for the sampling error influencing the present study's data and, consequently, will generally be less conservative than estimates of the effect in future samples.

Table 1. Various Adjusted R^2 Formulae.

Population Effect Estimates		Future Sample Effect Estimates	
Index	Formula	Index	Formula
Smith	$1-\left(\frac{n}{n-k}\right)\left(1-R^2\right)$	Lord-1	$1-\frac{n+k+1}{n-k-1}\left(1-R^2\right)$
Ezekiel	$1-\left(\frac{n-1}{n-k-1}\right)\left(1-R^2\right)$	Lord-2	$1-\frac{(n+k+1)(n-1)}{(n-k-1)n}\left(1-R^2\right)$
Wherry-2	$1-\left(\frac{n-1}{n-k}\right)\left(1-R^2\right)$	Darlington-Stein	$1-\left(\frac{n-1}{n-k-1}\right)\left(\frac{n-2}{n-k-2}\right)\left(\frac{n+1}{n}\right)\left(1-R^2\right)$
Olkin-Pratt	$R^2-\frac{k-2}{n-k-1}\left(1-R^2\right)-\left(\frac{2(n-3)}{(n-k-1)(n-k+1)}\right)\left(1-R^2\right)^2$	Browne[a]	$\frac{(n-k-3)\rho^4+\rho^2}{(n-2k-2)\rho^2+\rho}$
Pratt	$1-\frac{(n-3)(1-R^2)}{(n-k-1)}\left[1+\frac{2(1-R^2)}{n-k-2.3}\right]$	Claudy-1[b]	$\left(2\rho-R\right)^2$
Claudy-3	$1-\frac{(n-4)\left(1-R^2\right)}{(n-k-1)}\left[1+\frac{2\left(1-R^2\right)}{n-k+1}\right]$	Claudy-2	$1-\left(\frac{n-1}{n-k-1}\right)\left(\frac{n-2}{n-k-2}\right)\left(\frac{n-1}{n}\right)\left(1-R^2\right)$
		Rozeboom-1	$1-\left(\frac{n+k}{n-k}\right)\left(1-R^2\right)$
		Rozeboom-2[a]	$\rho^2\left[1+\left(\frac{k}{n-k-2}\right)\left(\frac{1-\rho^2}{\rho^2}\right)\right]^{-1}$

Note. n=sample size. k=number of predictor variables. Adapted from Yin & Fan (2001) and, Snyder & Lawson (1993). [a]ρ2 was estimated with the Ezekiel value. [b]ρ was estimated with the square root of the Ezekiel value. Negative Ezekiel values were replaced with zeros.

Ultimately, the decision of which R^2 adjustment formula to use depends on the generalizations that the researcher wishes to make. As Snyder and Lawson (1993) observed, "Most researchers ground their work in empirical findings from previous samples and usually desire that their work generalize to future samples" (p. 341). Researchers seeking this goal would be wise to consider corrected effect size estimates for future samples. If, however, the researcher wishes to develop population expectations, a population effect estimate may be more appropriate (Snyder & Lawson, 1993; Yin & Fan, 2001). If replicability is indeed the hallmark of scientific inquiry, then the sample effect that best represents the effect expected in the population or future samples should be of primary analytic interest. Accordingly, we argue that these corrected effects should be both reported and interpreted whenever possible.

Method

We examined regression applications in four journals –*Journal of Applied Psychology* (v.86[5] – 87[4]), *Journal of Educational Psychology* (v. 93[4]-94[3]), *Journal of Experimental Education* (v. 95[1]-96[1]), *and Journal of Educational Research* (v. 70[2]-71[1])– over a one-year time span. We considered only the first three regression analyses presented in each article. Additional analyses were not considered so that articles containing an above-average number of regression analyses would not overly impact the results. The frequencies of uncorrected and corrected effects were coded as well as the interpretation of the effects. We considered an effect to be interpreted if the author included a statement explaining the effect in relation to the dependent variable. For example, Klein, Conn, and Sorra (2001) interpreted R^2 by noting, "Together management support and financial resource availability explained 19%. . .of the variance in implementation of polices and practices (β = .36, p < .05)" (p. 819). This and similar statements were coded as interpreted.

Table 2. Reporting and Interpretation Frequency of Uncorrected and Corrected Effect Sizes.

Journal	No. of Articles Using Mult. Regression	No. of Mult. Regression Analyses	No. not Reporting an Effect Size	No. Reported	No. Interpreted	No. Reported	No. Interpreted
Journal of Applied Psychology	9	22	0 (0.00)	16 (72.73)	7 (31.82)	9 (39.13)	0 (0.00)
Journal of Educational Psychology	11	28	12 (42.85)	15 (53.57)	9 (32.14)	1 (3.57)	1 (3.57)
Journal of Educational Research	4	9	3 (33.33)	4 (44.44)	3 (33.33)	3 (33.33)	2 (22.22)
Journal of Experimental Education	1	2	0 (0.00)	2 (100.00)	0 (0.00)	0 (0.00)	0 (0.00)
Total	25	61	15 (24.59)	37 (60.65)	19 (31.15)	13 (20.97)	3 (4.92)

Note. The first, second, and third multiple regression analyses were considered from each article. Percentages are presented in parentheses under selected frequencies. The number of uncorrected and corrected effects may sum to greater than the total number of analyses due to the fact that some analyses reported both types of effects.

Results

Reporting Frequency

Reporting frequencies of uncorrected and corrected effects are displayed in Table 2. Overall, 61% of the analyses reported an uncorrected effect size. Fewer interpreted the effects, however, numbering roughly 50% of the total uncorrected effects reported. These results are relatively consistent with previous studies' findings addressing uncorrected effect sizes and their interpretation (cf. Henson & Smith, 2000; Kirk, 1996; Thompson & Snyder, 1997; Vacha-Haase, Nilsson, Reetz, Lance & Thompson, 2000).

Corrected effects occurred much less often in the literature, showing up in only 21% of the reviewed articles. Interpretation of the corrected effects was even rarer at 5% of all adjusted effects reported. Adjusted R^2 was reported more frequently than other types of corrected effects to the near exclusion of other options (ω^2 was reported in one instance), but the formulae used to calculate adjusted R^2 were not reported in the literature. Nevertheless, one could reasonably surmise that most of the authors likely used the Ezekiel formula (sometimes incorrectly attributed to Wherry [Yin & Fan, 2001]) because it is the formula used by the popular SAS and SPSS statistical software packages to calculate adjusted R^2 (Kirk, 1996; Yin & Fan, 2001). Although use of the Ezekiel correction is better than no correction, the near complete dependence on it in the present review begs the issues of (a) whether authors are reporting adjusted R^2 by default because it is provided in statistical output and (b) whether authors are aware of other correction options.

Adjusted R^2 using Various Formulae

For comparative purposes, we calculated adjusted effect sizes for all analyses that included an uncorrected effect size. We used information provided by the journal authors to adjust R^2 using each of the formulae listed in Table 1. Tables 3 and 4 present the uncorrected R^2, followed by the values calculated for the population and future sample adjustment formulae, respectively. These calculations demonstrate the amount of correction for each formula in relation to the uncorrected effect size, number of predictors, and sample size. Tables 5, 6, and 7 present the degrees of shrinkage for each of the adjustment formulae categorized by sample size, uncorrected R^2, and number of predictors, respectively.

Discussion

Reporting and Interpretation Frequency of Adjusted Effects

As noted, the reporting and interpretation of adjusted effects was rare. In fact, only 3 of the 62 (4.92%) regression analyses reviewed in this study reported and interpreted an adjusted effect. Given the

Table 3. Adjusted R2 Using Population Effect Adjustment Formulae

N	*k*	Reported Adj. R^2	R^2	Smith	Ezekiel	Wherry-2	Olkin-Pratt	Pratt	Claudy-3
578	2	-	.01	.0066	.0066	.0083	.0066	.0066	.0083
1340	3	-	.02	.0178	.0178	.0185	.0178	.0178	.0186
473	1	-	.03	.0279	.0279	.0300	.0281	.0281	.0302
1261	4	-	.03	.0269	.0269	.0277	.0270	.0270	.0277
99	2	-	.05	.0304	.0302	.0402	.0316	.0309	.0417
463	8	-	.12	.1045	.1045	.1065	.1049	.1049	.1069
463	8	-	.12	.1045	.1045	.1065	.1049	.1049	.1069
465	1	-	.13	.1281	.1281	.1300	.1286	.1286	.1305
62	1	-	.14	.1259	.1257	.1400	.1309	.1296	.1456
465	1	-	.14	.1381	.1381	.1400	.1387	.1387	.1405
465	1	-	.14	.1381	.1381	.1400	.1387	.1387	.1405
1515	6	-	.14	.1366	.1366	.1372	.1367	.1367	.1373
1515	6	-	.14	.1366	.1366	.1372	.1367	.1367	.1373
1515	6	-	.14	.1366	.1366	.1372	.1367	.1367	.1373
99	2	-	.17	.1529	.1527	.1614	.1559	.1555	.1647
343	10	-	.17	.1451	.1450	.1476	.1458	.1457	.1483
664	12	-	.18	.1649	.1649	.1662	.1653	.1653	.1666
37	2	-	.19	.1437	.1424	.1669	.1536	.1499	.1784
463	8	-	.20	.1859	.1859	.1877	.1866	.1866	.1884
187	11	-	.22	.1713	.1710	.1757	.1727	.1725	.1772
62	3	-	.24	.2014	.2007	.2142	.2073	.2062	.2207
99	3	-	.26	.2369	.2366	.2446	.2408	.2404	.2487
170	8	-	.26	.2235	.2232	.2280	.2255	.2253	.2301
24	3	-	.29	.1886	.1835	.2224	.2064	.1979	.2442
36	3	-	.37	.3127	.3109	.3318	.3262	.3236	.3467
45	3	-	.42	.3786	.3776	.3924	.3898	.3885	.4044
35	4	-	.50	.4355	.4333	.4516	.4500	.4481	.4672
289	7	.54	.55	.5388	.5388	.5404	.5405	.5405	.5421
412	7	-	.59	.5829	.5829	.5839	.5841	.5841	.5851
289	7	.61	.62	.6106	.6105	.6119	.6122	.6122	.6136
25	3	-	.63	.5795	.5771	.5964	.5999	.5978	.6181
170	8	-	.64	.6222	.6221	.6244	.6249	.6249	.6272
288	8	.65	.66	.6503	.6503	.6515	.6518	.6518	.6531
25	3	-	.67	.6250	.6229	.6400	.6444	.6427	.6605
35	4	-	.74	.7065	.7053	.7148	.7182	.7176	.7270
35	3	-	.75	.7266	.7258	.7344	.7380	.7376	.7462
38	4	-	.80	.7765	.7758	.7824	.7855	.7852	.7916
	M	.54	.31	.2870	.2864	.2938	.2917	.2910	.2989
	SD	.06	.24	.2382	.2379	.2392	.2414	.2413	.2426

Note. Population adjusted effect estimates were calculated for analyses that reported uncorrected R^2 values. Analyses that did not include an uncorrected effect size were not included. *n*=sample size. *k*=number of predictor variables.

fact that adjusted effects theoretically provide the researcher with a more realistic picture of the treatment effect, this result is surprisingly low.

Comparison of Various Adjustment Formulae

For those analyses not reporting an adjusted effect, we calculated and compared adjustments using the each of the fourteen adjustment formulae. Of the adjusted R^2 formulae estimating the population effect, the Ezekiel formula provided the most conservative correction for sampling error while the

Table 4. Adjusted R2 Using Future Sample Effect Adjustment Formulae.

N	*k*	Reported Adj. R^2	R^2	Lord-1	Lord-2	Darlington -Stein	Browne	Claudy-1	Claudy-2	Rozeboom -1	Rozeboom -2
578	2	-	.01	-.0003	.0014	.0014	.0081	.0038	.0048	.0031	.0043
1340	3	-	.02	.0141	.0149	.0149	.0185	.0157	.0163	.0156	.0158
473	1	-	.03	.0218	.0238	.0238	.0297	.0260	.0279	.0259	.0260
1261	4	-	.03	.0223	.0231	.0230	.0276	.0240	.0246	.0238	.0241
99	2	-	.05	-.0094	.0008	-.0002	.0389	.0154	.0198	.0108	.0180
463	8	-	.12	.0851	.0871	.0867	.1076	.0901	.0906	.0891	.0908
463	8	-	.12	.0851	.0871	.0867	.1076	.0901	.0906	.0891	.0908
465	1	-	.13	.1225	.1244	.1244	.1295	.1263	.1281	.1263	.1263
62	1	-	.14	.0827	.0975	.0965	.1363	.1121	.1252	.1118	.1124
465	1	-	.14	.1326	.1344	.1344	.1395	.1363	.1381	.1363	.1363
465	1	-	.14	.1326	.1344	.1344	.1395	.1363	.1381	.1363	.1363
1515	6	-	.14	.1320	.1326	.1326	.1375	.1332	.1337	.1332	.1332
1515	6	-	.14	.1320	.1326	.1326	.1375	.1332	.1337	.1332	.1332
1515	6	-	.14	.1320	.1326	.1326	.1375	.1332	.1337	.1332	.1332
99	2	-	.17	.1181	.1270	.1261	.1607	.1363	.1436	.1358	.1367
343	10	-	.17	.1150	.1176	.1166	.1509	.1220	.1217	.1202	.1231
664	12	-	.18	.1473	.1485	.1482	.1686	.1504	.1507	.1498	.1508
37	2	-	.19	.0471	.0728	.0658	.1651	.1016	.1150	.0974	.1043
463	8	-	.20	.1683	.1701	.1697	.1901	.1723	.1733	.1719	.1726
187	11	-	.22	.1130	.1178	.1138	.1849	.1281	.1233	.1225	.1309
62	3	-	.24	.1352	.1491	.1451	.2176	.1649	.1722	.1627	.1659
99	3	-	.26	.1977	.2058	.2043	.2472	.2144	.2202	.2138	.2145
170	8	-	.26	.1773	.1821	.1796	.2368	.1893	.1892	.1869	.1901
24	3	-	.29	.0060	.0474	.0152	.2347	.1013	.0940	.0871	.1078
36	3	-	.37	.2125	.2344	.2233	.3467	.2570	.2652	.2555	.2560
45	3	-	.42	.3068	.3222	.3160	.4071	.3374	.3457	.3371	.3360
35	4	-	.50	.3333	.3524	.3367	.4952	.3714	.3736	.3710	.3671
289	7	.54	.55	.5244	.5260	.5256	.5515	.5277	.5289	.5277	.5275
412	7	-	.59	.5738	.5748	.5746	.5923	.5758	.5767	.5758	.5757
289	7	.61	.62	.5984	.5998	.5994	.6247	.6011	.6022	.6011	.6009
25	3	-	.63	.4890	.5095	.4943	.6533	.5266	.5332	.5291	.5200
170	8	-	.64	.5998	.6021	.6009	.6519	.6045	.6056	.6044	.6038
288	8	.65	.66	.6381	.6393	.6389	.6677	.6406	.6414	.6406	.6403
25	3	-	.67	.5443	.5625	.5489	.7026	.5774	.5836	.5800	.5710
35	4	-	.74	.6533	.6632	.6551	.7922	.6715	.6743	.6729	.6669
35	3	-	.75	.6855	.6945	.6898	.7826	.7020	.7070	.7031	.6994
38	4	-	.80	.7394	.7463	.7411	.8597	.7519	.7544	.7529	.7487
	M	.54	.31								
	SD	.06	.24	.2489	.2565	.2528	.3076	.2649	.2676	.2640	.2646
				.2341	.2340	.2331	.2581	.2332	.2335	.2343	.2316

Note. Future sample adjusted effect estimates were calculated for analyses that reported uncorrected R^2 values. *n*=sample size. *k*=no. of predictor variables.

Table 5. Degree of Shrinkage Categorized by Number of Predictors

Population Effect Estimates												
	Smith		Ezekiel		Wherry-2		Olkin-Pratt		Pratt		Claudy-3	
K	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1-2[a]	.0120	.0148	.0122	.0152	.0048	.0079	.0097	.0118	.0104	.0130	.0022	.0053
3-4[b]	.0391	.0266	.0404	.0279	.0267	.0180	.0284	.0218	.0300	.0239	.0153	.0127
5-7[c]	.0063	.0034	.0063	.0035	.0054	.0030	.0055	.0027	.0055	.0027	.0045	.0022
8-9[d]	.0182	.0094	.0183	.0095	.0159	.0082	.0169	.0091	.0169	.0091	.0146	.0079
10+[e]	.0296	.0173	.0297	.0174	.0269	.0157	.0288	.0168	.0288	.0169	.0260	.0152

Future Sample Effect Estimates																
	Lord-1		Lord-2		Darlington-Stein		Browne		Claudy-1		Claudy-2		Rozeboom-1		Rozeboom-2	
K	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1-2	.0392	.0454	.0315	.0372	.0326	.0394	.0059	.0082	.0229	.0281	.0177	.0242	.0240	.0295	.0222	.0272
3-4	.1062	.0738	.0919	.0632	.1010	.0717	-.0050	.0330	.0773	.0504	.0735	.0516	.0781	.0531	.0790	.0499
5-7	.0146	.0078	.0136	.0073	.0138	.0075	-.0001	.0031	.0126	.0069	.0119	.0066	.0126	.0069	.0127	.0070
8-9	.0411	.0213	.0387	.0200	.0396	.0209	.0064	.0134	.0355	.0180	.0349	.0184	.0363	.0188	.0353	.0177
10+	.0649	.0381	.0620	.0363	.0638	.0382	.0219	.0121	.0565	.0320	.0581	.0348	.0592	.0346	.0551	.0308

Note. N = 37 analyses. Shrinkage = uncorrected R^2 – adjusted R^2. [a]n = 9. [b]n = 13. [c]n = 6. [d]n = 6. [e]n=3.

Table 6. Degree of Shrinkage Categorized by Sample Size.

Population Effect Estimates												
Study	Smith		Ezekiel		Wherry-2		Olkin-Pratt		Pratt		Claudy-3	
N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1-30[a]	.0656	.0311	.0688	.0327	.0438	.0207	.0464	.0323	.0505	.0361	.0224	.0203
31-50[b]	.0414	.0159	.0427	.0164	.0280	.0116	.0298	.0145	.0314	.0155	.0155	.0097
51-100[c]	.0225	.0096	.0228	.0098	.0119	.0095	.0187	.0088	.0195	.0088	.0077	.0091
100+[d]	.0115	.0121	.0115	.0122	.0097	.0113	.0071	.0068	.0107	.0118	.0089	.0109

Future Sample Effect Estimates																
Study	Lord-1		Lord-2		Darlington-Stein		Browne		Claudy-1		Claudy-2		Rozeboom-1		Rozeboom-2	
N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1-30	.1836	.0873	.1569	.0745	.1772	.0848	-.0002	.0483	.1282	.0527	.1264	.0605	.1313	.0622	.1304	.0452
31-50	.1132	.0439	.0977	.0375	.1060	.0409	.0112	.0361	.0825	.0307	.0764	.0302	.0829	.0318	.0845	.0305
51-100	.0691	.0214	.0560	.0201	.0576	.0215	.0119	.0068	.0434	.0189	.0358	.0200	.0450	.0192	.0425	.0189
100+	.0265	.0263	.0247	.0253	.0252	.0262	.0050	.0106	.0147	.0135	.0217	.0242	.0229	.0243	.0220	.0226

Note. N = 37 analyses. Shrinkage = uncorrected R^2 – adjusted R^2. [a]n = 3. [b]n = 7. [c]n = 5. [d]n = 22.

Table 7. Degree of Shrinkage Categorized by Uncorrected R^2.

Population Effect Estimates

	Smith		Ezekiel		Wherry-2		Olkin-Pratt		Pratt		Claudy-3	
R^2	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
.01-.15[a]	.0065	.0065	.0066	.0065	.0036	.0049	.0059	.0060	.0060	.0061	.0029	.0053
.16-.30[b]	.0366	.0261	.0374	.0275	.0205	.0179	.0320	.0214	.0335	.0237	.0213	.0140
.31-.50[c]	.0544	.0018	.0561	.0124	.0381	.0104	.0413	.0101	.0433	.0106	.0239	.0086
.51+[d]	.0231	.0154	.0239	.0163	.0170	.0097	.0150	.0082	.0156	.0089	.0086	.0032

Future Sample Effect Estimates

	Lord-1		Lord-2		Darlington-Stein		Browne		Claudy-1		Claudy-2		Rozeboom-1		Rozeboom-2	
R^2	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
.01-.15	.0189	.0193	.0160	.0159	.0162	.0162	.0039	.0045	.0125	.0120	.0103	.0110	.0130	.0129	.0129	.0115
.16-.30	.0955	.0751	.0842	.0632	.0896	.0725	.0223	.0142	.0699	.0480	.0677	.0510	.0732	.0521	.0521	.0461
.31-.50	.1458	.0286	.1270	.0260	.1380	.0306	.0137	.0093	.1081	.0234	.1018	.0262	.1088	.0236	.0236	.0247
.51+	.0604	.0447	.0532	.0376	.0581	.0432	-.0229	.0210	.0047	.0318	.0443	.0297	.0462	.0308	.0308	.0344

Note. N = 37 analyses. Shrinkage = uncorrected R^2 – adjusted R^2. [a]n = 14. [b]n = 10. [c]n = 3. [d]n = 10.

Table 8. *Selected Study R^2 Adjustments*

			Population Effect Estimates							
Study *n*	*k*	R^2	Smith	Ezekiel	Wherry-2	Olkin-Pratt	Pratt	Claudy-3		
38	4	.80[a]	.7765	.7758	.7824	.7855	.7852	.7916		
45	3	.42[b]	.3786	.3776	.3924	.3898	.3885	.4044		
25	3	.63[c]	.5795	.5771	.5964	.5999	.5978	.6181		
463	8	.12[d]	.1045	.1045	.1065	.1049	.1049	.1069		
24	3	.29[e]	.1886	.1835	.2224	.2064	.1979	.2442		
			Future Sample Effect Estimates							
Study *n*	*k*	R^2	Lord-1	Lord-2	Darlington-Stein	Browne	Claudy-1	Claudy-2	Rozeboom-1	Rozeboom-2
38	4	.80	.7394	.7463	.7411	.8597	.7519	.7544	.7529	.7487
45	3	.42	.3068	.3222	.3160	.4071	.3374	.3457	.3371	.3360
25	3	.63	.4890	.5095	.4943	.6533	.5266	.5332	.5291	.5200
463	8	.12	.0851	.0871	.0867	.1076	.0901	.0906	.0891	.0908
24	3	.29	.0060	.0474	.0152	.2347	.1013	.0940	.0871	.1078

Note: [a]Roth, Speece, & Cooper (2002); [b]Marks, Sabella, Burke, & Zaccaro (2002); [c]Gefland, Nishii, Holcombe, Dyer, Ohbuchi, & Fukuno (2001);, [d]Harackiewicz, Barron, Tauer, & Elliot (2002).

Claudy-3 formula offered the least conservative correction. Table 3 illustrates this trend for these adjustments. One can infer that most adjusted R^2 effects presented in the literature offer a conservative estimate since the Ezekiel correction is used in the SAS and SPSS (Kirk, 1996) software packages commonly used by researchers. The uninitiated researcher may not know, however, that these software packages use a formula that estimates only the population parameter.

As illustrated in Table 4, the majority of the future sample effect estimates provided even more conservative estimates than those predicting the population effect. Of these adjustment formulae, the Lord-1 provided the most conservative estimate of the future sample effect while the Browne formula provided the most liberal overall. It is important to note that nine of the 62 adjustments using the Browne formula actually resulted in corrected effects that were greater than the uncorrected effects – a logically impossible result. That is, a sample cannot possess less sampling error than a population that, by definition, has no sampling error at all. This phenomenon with the Browne formula begs further investigation. It would seem prudent to use caution with the Browne formulae for correction of effects of greater magnitude.

In shrinking R^2, adjustment formulae consider the three factors that affect sampling error: (a) sample size, (b) number of variables in the model, and the (c) uncorrected effect size (as an estimate of the population effect). It naturally follows that these three factors would affect the degree of correction provided by the adjustments to R^2.

Table 5 provides evidence of the number of predictor's impact on the degree of shrinkage. Generally, as the number of predictors increased, the degree of shrinkage increased as well. Our results may be somewhat inconsistent, however, as the analyses with 5-7 predictors did not always show the upward trend as expected. This may be due to the fact that the majority of the analyses in this group had large sample sizes. In fact, no analysis in this group reported a sample size less than 289 subjects. Consequently, this group may not be representative of the adjustment trend based on the number of predictors typically found in the research literature; the large sample sizes may have lessened the degree of correction for this group.

Thompson and Melancon (1990) reported that "with a very large effect size, or a large sample size, or both, it will matter less which, if any, statistical corrections the researcher applies in estimating effect sizes" (p. 11). This proposition is supported by the data in Table 6. As sample size increased, the amount of correction lessened, although in varying amounts based on the formulae. This is logical given that as sample size increases, sampling error – the issue for which adjustments are made – decreases. More specific evidence of this fact is provided in Table 7. Given the case with a large sample size (n = 463) and a small effect size ($R^2 = 0.12$), the correction was relatively small.

Thompson and Melancon (1990) also noted the converse – that statistical corrections tend to be greater when effect sizes and sample sizes are small. This can be noted generally in Tables 6 and 7. As sample size decreased, adjustments generally increased. Again, it is interesting to note that a smaller sample size typically results in greater theoretical sampling error in a sample. Because adjustments to R^2 are determined by the degree of sampling error, it follows that one could expect a large correction given a small sample size.

The correction for one case detailed in Table 8 provides specific evidence for this proposition. With a small sample size (n=24) and a moderate effect size (R^2=0.29), the adjustment was relatively large. Moreover, our results indicate that, in this case, result interpretation may have been different if the author had calculated adjusted R^2. The uncorrected R^2 presented in the journal article indicated that the treatment explained 29% of the dependent variable variance. When corrected using the Lord-1 and Darlington-Stein formulae, however, R^2 shrunk to near zero (.0060 and .0152, respectively.) In other words, after having corrected the effect size for sampling error expected upon replication of the study with a new sample, the treatment accounted for virtually no variance in the dependent variable, a fact that was obscured by the uncorrected R^2.

As demonstrated by the previous case, it is quite possible to overestimate the importance of a result if effect sizes are not adjusted to account for the influence of sampling error. Accordingly, researchers should report and interpret corrected effect measures in their results. Not only do corrected effects provide a better estimate of the effect in the population, they can provide information concerning the replicability of the results. When a researcher uses corrected effect sizes, we recommend that he or she take into account the various formulae, their purposes, and their relative degrees of correction. Such choices have the potential to directly impact results and their interpretation.

References

*(References marked with an asterisk indicate studies included in review)

American Psychological Association. (1994). Publication Manual of the American Psychological Association (4th ed.). Washington, DC: Author.

American Psychological Association. (2001). Publication Manual of the American Psychological Association (5th ed.). Washington, DC: Author.

*Bartol, K. M., Durham, C. C., Poon, J. M. L. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. Journal of Applied Psychology, 86, 1103–1119.

Berkson, J. (1942). Tests of signficance considered as evidence. Journal of American Statistical Association, 37, 325-335.

*Braaksma, M. A. H., Riglaarsdam, G., van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. Journal of Educational Psychology, 94, 405–415.

*Broekkamp, H., van Hout-Wolters, B. H. A. M., Rijlaarsdam, G., & van den Bergh, H. (2002). Importance in instructional text: Teachers' and students' perceptions of task demands. Journal of Educational Psychology, 94, 260–271.

*Cappella, E., Weinstein, R. S. (2001). Turning around reading achievement: Predictors of high school students' academic resilience. Journal of Educational Psychology, 93, 758–771.

Carter, D.S. (1979). Comparison of different shrinkage formulas in estimating the population multiple correlation coefficients. Educational and Psychological Measurement, 39, 261-266.

Cohen, J. (1994). The earth is round (p<.05). American Psychologist, 49, 997-1003.

Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. The Journal of Educational Research, 94, 275–282.

Finch, S., Cumming, G., Thomason, N. (2001). Reporting of statistical inference in the Journal of Applied Pscyhology: Little evidence of reform. Educational and Psychological Measurement, 61, 181-210.

*Gelfand, M. J., Nishii, L. H., Holcombe, K. M., Dyer, N., Ohbuchi, K., & Fukuno, M. (2001). Cultural influences on cognitive representations of conflict: Interpretations of conflict episodes in the United States and Japan. Journal of Applied Psychology, 86, 1059–1074.

*Gentry, M., Gable, R. K., & Rizza, M. G. (2002). Students' perceptions of classroom activities: Are there grade-level and gender differences? Journal of Educational Psychology, 94, 539–544.

*Griffiths, Y. M. & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. Journal of Educational Psychology, 94, 34–43.

*Harackiewica, J. M., Barron, K. E., Tauer, J. M., & Elliot, A.J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. Journal of Educational Psychology, 94, 562–575.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). What if there were no significance tests? Mahwah, NJ: Erlbaum.

*Hausknecht, J. P., Trevor, C. O., & Farr, J. L. (2002). Retaking Ability Tests in a Selection setting: Implications for practice effects, training performance, and turnover. Journal of Applied Psychology, 87, 243–254.

*Helwig, R., Anderson, L., & Tindal, G. (2001). Influence of elementary student gender on teachers' perceptions of mathematics achievement. The Journal of Educational Research, 95, 93–102.

Henson, R. K. (2006). Effect size measures and meta-analytic thinking in counseling
psychology research. The Counseling Psychologist, 34, 601-629.

Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force Report and current trends. Journal of Research and Development in Education, 33, 285-296.

*Hill, N. E. (2001). Parenting and academic socialization as they relate to school readiness: The roles of ethnicity and family income. Journal of Educational Psychology, 93, 686–697.

*Hinds, P. J., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. Journal of Applied Psychology, 86, 1232–1243.

Huberty, C. J. (1987). On statistical testing. Educational Researcher, 16, 4-9.

Huberty, C. J. (2002). A history of effect size indices. Educational and Psychological Measurement, 62, 227-240.

Huberty, C. J., & Pike, C. J. (1999). On some history regarding statistical testing. In B. Thompson (Ed.), Advances in social science methodology (Vol. 5, pp. 1-22). Stamford, CT: JAI Press.

*Jockin, V., Arvey, R. D., & McGue, M. (2001). Perceived victimization moderates self-reports of workplace aggression and conflict. Journal of Applied Psychology, 86, 1262–1269.

*Johnstone, K. M., Ashbaugh, H., & Warfield, T. D. (2002). Effects of repeated practice and contextual-writing experiences on college students' writing skills. Journal of Educational Psychology, 94, 305–315.

Kirk, R. E. (1996). Practical significance: A concept whose time has come. Educational and Psychological Measurement, 56, 746–759.

*Kitsantas, A. (2002). Test preparation and performance: A self-regulatory analysis. The Journal of Experimental Education, 70, 101-113.

*Klein, K. J., Conn, A. B., & Sorra, J. S. (2001). Implementing computerized technology: An organizational analysis. Journal of Applied Psychology, 86, 811–824.

Kline, R.B. (2004). Beyond significance testing: Reforming data analysis methods in behavioral research. Washington, D.C.: American Psychological Association.

Larson, S.C. (1931). The shrinkage of the coefficient of multiple correlation. Journal of Educational Psychology, 22, 45-55.

*Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. Journal of Applied Psychology, 87, 3–13.

*McGregor, H. A. & Elliot, A. J. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. Journal of Educational Psychology, 94, 381–395.

*Miller, D. C. & Byrnes, J. P. (2001). To achieve or not to achieve: A self-regulation perspective on adolescents' academic decision making. Journal of Educational Psychology, 93, 677–685.

*Okpala, C. O., Okpala, A. O., & Smith, F. E. (2001). Parental involvement, instructional expenditures, family socioeconomic attributes, and student achievement. The Journal of Educational Research, 95, 110–115.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparitive studies: Applications, interpretations, and limitations. Contemporary Educational Psychology, 25, 241-286.

Onwuegbuzie, A.J., Levin, J.R., & Leech, N.L. (2003). Do effect-size measures measure up?: A brief assessment. Learning disabilities: A contemporary journal, 1(1): 37-40.

Pedhazur, E. J. (1997). Multiple Regression in Behavioral Research (3rd ed.). Fort Worth: Harcourt Brace.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillside, NJ: Erlbaum.

*Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. Journal of Educational Psychology, 94, 396–404.

Roberts, J. K., & Henson, R. K. (2002). Correction for bias in estimating effect sizes. Educational and Pscyhological Measurement, 62, 241-253.

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276–1284.

*Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. The Journal of Educational Research, 95, 259–272.

*Schleicher, D. J., Day, D. V., Mayes, B. T., Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. Journal of Applied Psychology, 87, 735–746.

*Settles, I. H., Sellers, R. M., & Damas, A., Jr. (2002). One role or two? The function of psychological separation in role conflict. Journal of Applied Psychology, 87, 574–582.

Snyder, P. & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. Journal of Experimental Education, 61, 334–349.

Thompson, B. (1994). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62, 157–176.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25, 26–30.

Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. Educational Researcher, 26, 29–32.

Thompson, B. (1999). If statistical significance tests are broken/misused, what practices should supplement or replace them [Electronic version]? Theory & Psychology, 9, 165–181.

Thompson, B. (2002). “Statistical,” “practical,” and “clinical”: How many kinds of significance do counselors need to consider? Journal of Counseling and Development, 80(1), 64-72.

Thompson, B., & Kieffer, K. M. (2000). Interpreting statistical significance test results: A proposed new "what if" method. Research in the Schools, 7(2), 3–10.

Thompson, B., & Melancon, J. G. (1990, November). Bootstrap versus statistical effect size corrections: A comparison with data from the finding embedded figures test. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans.

Thompson, B., & Snyder, P. A. (1997). Statistical significance testing practices in the Journal of Experimental Education [Electronic version]. Journal of Experimental Education, 66, 75–83.

Tyler, R. W. (1931). What is statistical significance? Educational Research Bulletin, 10, 115-118, 142.

Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding statistical significance and effect size. Theory & Psychology, 10, 413-425.

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. Journal of Counseling Psychology, 51, 473-481.

*Wilkins, J. L. M. & Ma, X. (2002). Predicting student growth in mathematical content knowledge. The Journal of Educational Research, 95, 288-298.

Wilkinson, L., & American Psychological Association (APA) Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations [Electronic version]. American Psychologist, 54, 594-604.

Yin, P. & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of analytical methods. The Journal of Experimental Education, 69, 203-224.

Send correspondence to: Robin K. Henson
University of North Texas
Email: rhenson@unt.edu