Male and female differences in self-report cheating

James A Athanasou
University of Technology, Sydney

Olabisi Olasehinde
University of Ilorin – Nigeria

Cheating is an important area for educational research, not only because it reduces the consequential validity of assessment results, but also because it is anathema to widely held public principles of equity and truthfulness (see Cizek, 1999 for a comprehensive review of the topic). Moreover, modern education is centred on numerous situations that really depend upon a student’s honesty. The purpose of this paper is to review the extent of academic cheating and to describe any gender differences in self-reports.

There is a large literature relating to the influence of gender on academic cheating or dishonesty; the topic has captured the attention of researchers since the pioneering work of Hartshorne and May on deceit and honesty in 1928. It has been part of a wider fascination with the ethical, moral, and social dimensions of this academic behaviour at all levels of education.

In the course of various studies of academic honesty, it has been suggested that there are gender differences in the extent of cheating in education and that overall, women are less likely to cheat, but this conclusion has been challenged (Black, 1962; Graham, Monday, O’Brien & Steffen, 1994; Hartshorne & May, 1928; Kerkvliett, 1994; McCabe & Trevino, 1996). Good, Nichols, and Sabers (1999) noted that there may be differences in the perception of cheating between males and females, yet a meta-analytic study of gender and sex roles in relation to cheating produced a low mean effect size of 0.19 for self-reports of cheating (Whitley, Nelson & Jones, 1999). This paper considers some key issues in past research on cheating and reviews those studies that have investigated the influence of gender on self-reported cheating.

Academic cheating

There is no consensus in estimates of the extent of cheating but it has been viewed as a major problem, with the majority of students indicating that they have been dishonest (Baird, 1980; Sierles, Hendrickx & Circle, 1980; Whitley, 1998). When specific forms of cheating such as plagiarism, collusion, copying, etc., are investigated, then the proportions reporting that they have been dishonest are reduced (see Hollinger & Lanza-Kaduce, p. 293). Kerkvliett and Sigmund (1999) discussed the prevalence of cheating within university and college systems and concluded: “The evidence indicates that many students cheat regularly and few students never cheat” (p. 331).

At the outset, it may be helpful to describe cheating for the reader as it includes a variety of behaviours. The essence of cheating is fraud and deception. We have adapted a working description of cheating in education contexts from a discussion on academic dishonesty in nursing students (Gabertson, 1997), as involving conscious participation in deception (through lying, dishonesty, falsifying, misrepresenting, corruption, plagiarism, copying, or unlawfully assisting someone else). Newstead, Franklyn-Stokes and Armstead (1996) provided the following list of cheating behaviours (see also Baird, 1980; Franklyn-Stokes and Newstead, 1995, p. 164), which we have classified according to Cizek (1999, p. 39):

Cheating by taking, giving, or receiving information from others

- Allowing own coursework to be copied by another student
- Copying another student’s coursework with their knowledge
- Submitting a piece of coursework as an individual piece of work when it has actually been written jointly with another student
- Doing another student’s coursework for them
- Copying from a neighbour during an examination without them realising
- Copying another student’s coursework without their knowledge
- Submitting coursework from an outside source (e.g., a former student offers to sell pre-prepared essays, “essay banks”)
- Premeditated collusion between two or more students to communicate answers to each other during an
Hollinger and Lanza-Kaduce (1996) Despite its many limitations, the method of confidential and anonymous categories (see Chaudhuri & Mukerjee, 1988). Truthful responses can be determined based on comparisons of the assumed distribution of numbers and responses in the social security ID and this was categorised before they were required to answer truthfully. The probability of admitting to cheating using direct questioning (even when anonymous) was 0.259, compared with 0.419 using a randomized response technique. [In the randomized response technique, students generated a random number from an assigned distribution. The probability of cheating was then based on this number.]

Doubts about the credibility of this method have been noted (Bushway & Nash, 1977, p. 629; Spiller & Crown, 1995, p. 764). One, a feature of the self-reports is that they have yielded higher response rates from females in a number of studies (see McCabe & Trevino, 1997, p. 386). Secondly, useable response rates for some surveys have varied (e.g., 90% - Franklyn-Stokes and Newstead, 1995; 65% - Erickson & Smith, 1974) depending upon the circumstances under which the data was collected. Thirdly, the use of direct questioning methods may underestimate current class-specific cheating. For instance, Kerkvliett (1994) found that the proportion of students admitting to cheating using direct questioning (even when anonymous) was 0.259, compared with 0.419 using a randomized response technique. [In the randomized response technique, students generated a random number from their social security ID and this was categorised before they were required to answer truthfully. The probability of truthful responses can be determined based on comparisons of the assumed distribution of numbers and responses in the categories (see Chaudhuri & Mukerjee, 1988).]

Despite its many limitations, the method of confidential and anonymous self-report has ethical and moral advantages. Hollinger and Lanza-Kaduce (1996) stated:

... a self-administered survey provides the best opportunity to obtain detailed information from students about their academic dishonesty. It also avoids the ethical problems associated with contriving temptations to

---

**Cheating through the use of forbidden materials or information**

- Paraphrasing material from another source without acknowledging the original author
- Inventing data (i.e., nonexistent results)
- Fabricating references or a bibliography
- Copying material for coursework from a book or other publication without acknowledging the source
- Altering data (e.g., adjusting data to obtain a significant result)
- Taking unauthorised material into an examination

**Cheating by circumventing the process of assessment**

- Taking an examination for someone else or having someone else take an examination
- Attempting to obtain special consideration by offering or receiving favours through bribery, seduction, corruption
- Lying about medical or other circumstances to get special consideration by examiners (e.g., to get a more lenient view of results; extra time to complete the exam; an extended deadline; or exemption)
- Deliberately mis-shelving books or journal articles in the library so that other students cannot find them or by cutting out the relevant article or chapter
- Coming to an agreement with another student to mark each other’s work more generously than it merits
- Illicitly gaining advance information about the contents of an examination paper
- Concealing teacher or professor errors
- Threats or blackmail or extortion

Consequently, cheating involves a wide range of behaviours. They can vary in their seriousness, execution, purpose, and social dimensions.

Early descriptions of honesty emphasised that it was situation-specific (Hartshorne & May, 1928). Later investigations of cheating behaviours have looked at relationships with factors such as culture, socialisation, field of study, extent of competitiveness and gender (Bowers, 1964; McCabe & Trevino, 1995, 1996). This led to the development of a two-factor theory of morality based on generalised traits and specific predictors (Burton, 1963). Others see cheating as deviant behaviour and explain it in terms of (a) deterrence theory, in which the probability and extent of punishment control behaviours; (b) rational choice theory, in which the probabilities of both rewards and punishments are included; (c) social bond theory, in which deviant behaviour is a result of the weakening of social bonds such as attachment, commitment, involvement and moral belief; and (d) social learning theory, in which deviant behaviour is reinforced in primary groups (see Michaels & Miethe, 1989).

A small number of studies involved observational or experimental findings. Observational studies usually involved (a) some form of surreptitious observation in which students have the opportunity to cheat; or (b) determining the overlap in errors of adjacent students in an exam with the overlap in errors of non-adjacent students; or (c) a randomised response technique which invites a binary response from a student (see Chaudhuri & Mukerjee, 1988 for a description of this method). Experimental methods involve manipulations, such as the examination of cheating on sex-appropriate tasks involving 11 very difficult and 4 easy questions (Lobel, 1993) or cheating under high- and low- risk conditions (Leming, 1980) or cheating by copying assignments across semesters (Karlins, Michaels, Frellinger & Walker, 1989).

Observational and experimental manipulations are not included in this review.

**Self-reports of cheating**

Given the sensitivity of dishonest behaviours, most recent studies of cheating have relied upon survey methods involving anonymous self-report. Doubts about the credibility of this method have been noted (Bushway & Nash, 1977, p. 629; Spiller & Crown, 1995, p. 764). One, a feature of the self-reports is that they have yielded higher response rates from females in a number of studies (see McCabe & Trevino, 1997, p. 386). Secondly, useable response rates for some surveys have varied (e.g., 90% - Franklyn-Stokes and Newstead, 1995; 65% - Erickson & Smith, 1974) depending upon the circumstances under which the data was collected. Thirdly, the use of direct questioning methods may underestimate current class-specific cheating. For instance, Kerkvliett (1994) found that the proportion of students admitting to cheating using direct questioning (even when anonymous) was 0.259, compared with 0.419 using a randomized response technique. [In the randomized response technique, students generated a random number from their social security ID and this was categorised before they were required to answer truthfully. The probability of truthful responses can be determined based on comparisons of the assumed distribution of numbers and responses in the categories (see Chaudhuri & Mukerjee, 1988).]

Despite its many limitations, the method of confidential and anonymous self-report has ethical and moral advantages. Hollinger and Lanza-Kaduce (1996) stated:

... a self-administered survey provides the best opportunity to obtain detailed information from students about their academic dishonesty. It also avoids the ethical problems associated with contriving temptations to...
cheat and then deceiving students about it. Further, a survey instrument can be used to collect information efficiently about different forms of academic dishonesty across a variety of contexts. Surveys also permit students to remain anonymous. In general, confidential self-report surveys about minor forms of deviance among conventionally socialized individuals have been judged to be methodologically valid and reliable… (1996, p. 394)

This study reviews and evaluates the extent of any gender differences in academic cheating behaviours based on previous studies that used self-report data. The main research question was whether males reported higher rates of cheating than females, and a secondary question was whether any gender difference was consistent across assessment contexts. A meta-analysis was used to accumulate the results of previous studies because it offered a better representation of the relationship between gender and self-reports of cheating than can be provided by any one study. This meta-analysis used the effect size statistic, d (Cohen, 1988).

**Literature search strategy**

Computer-based searches of Psychological Abstracts and the Educational Resources Information Center (ERIC) databases were conducted only for published studies relating to gender and cheating. This was supplemented by checking the references cited for any further studies not located by the computer-based search. Twenty-one studies that reported on gender and cheating were located. These studies are summarised in Tables 1 and 2. Table 1 summarises those 14 studies that cited the proportion of men and women cheating and Table 2 lists the 21 studies used for the determination of an effect size.

**TABLE 1: Proportion of male and female students who reported cheating**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Sample</th>
<th>Females</th>
<th>Males</th>
<th>Country</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ameen, Guffey &amp; McMillan</td>
<td>1996</td>
<td>University</td>
<td>168</td>
<td>117</td>
<td>USA</td>
<td>0.518</td>
<td>0.624</td>
</tr>
<tr>
<td>Astin, Panos &amp; Creager</td>
<td>1967</td>
<td>College</td>
<td>94,537</td>
<td>112,328</td>
<td>USA</td>
<td>0.165</td>
<td>0.241</td>
</tr>
<tr>
<td>Baldwin, Daugherty, Rowley &amp; Schwarz</td>
<td>1996</td>
<td>Medical school</td>
<td>1,510</td>
<td>916</td>
<td>USA</td>
<td>0.302</td>
<td>0.468</td>
</tr>
<tr>
<td>Bowers</td>
<td>1964</td>
<td>University</td>
<td>2,568</td>
<td>2,810</td>
<td>USA</td>
<td>0.430</td>
<td>0.540</td>
</tr>
<tr>
<td>Burns, Davis, Hoshino &amp; Miller</td>
<td>1998</td>
<td>University</td>
<td>77</td>
<td>151</td>
<td>Japan</td>
<td>0.449</td>
<td>0.370</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>88</td>
<td>32</td>
<td>Sth Africa</td>
<td>0.193</td>
<td>0.438</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University</td>
<td>57</td>
<td>33</td>
<td>Sth Africa</td>
<td>0.404</td>
<td>0.545</td>
</tr>
<tr>
<td>Davis, Noble, Zak &amp; Dreyer</td>
<td>1994</td>
<td>University</td>
<td>39</td>
<td>10</td>
<td>Aust</td>
<td>0.510</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1,478</td>
<td>675</td>
<td>USA</td>
<td>0.760</td>
<td>0.790</td>
</tr>
<tr>
<td>Erickson &amp; Smith</td>
<td>1974</td>
<td>College students</td>
<td>68</td>
<td>50</td>
<td>USA</td>
<td>0.059</td>
<td>0.160</td>
</tr>
<tr>
<td>Huss, Curnyn, Roberts, Davis et al.</td>
<td>1993</td>
<td>College students</td>
<td>142</td>
<td>78</td>
<td>USA</td>
<td>0.730</td>
<td>0.770</td>
</tr>
<tr>
<td>Schab</td>
<td>1969</td>
<td>High school</td>
<td>580</td>
<td>835</td>
<td>USA</td>
<td>0.697</td>
<td>0.519</td>
</tr>
</tbody>
</table>
Schab 1980 High school 580 520 USA 0.952 0.899
Smith Ryan & Diggins 1972 College 68 44 USA 0.970 0.910
Thorpe, Pittenger & Reed 1999 University 81 57 USA 0.795 0.912

124 48 0.621 0.729

Who's who amongst American high school students 1993 High school 1,429 528 USA 0.679 0.660
Who's who amongst American high school students 1994 High school 2,256 921 USA 0.722 0.632

Note: Includes largest proportion where multiple proportions are quoted

TABLE 2: Effect sizes and meta-analysis

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Level</th>
<th>Female</th>
<th>Male</th>
<th>School</th>
<th>College</th>
<th>Quiz</th>
<th>Unit test</th>
<th>Mid-term</th>
<th>Exam</th>
<th>Copy</th>
<th>Plagiarism</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ameen, Guffey &amp; McMillan</td>
<td>1996</td>
<td>University</td>
<td>168</td>
<td>117</td>
<td>0.211</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antion &amp; Michael</td>
<td>1983</td>
<td>Community college</td>
<td>84</td>
<td>64</td>
<td>0.262</td>
<td>-0.242</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astin, Panos &amp; Creager</td>
<td>1967</td>
<td>Community college</td>
<td>94,537</td>
<td>112,328</td>
<td>0.188</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baird</td>
<td>1980</td>
<td>College</td>
<td>113</td>
<td>87</td>
<td>0.551</td>
<td>0.525</td>
<td>0.471</td>
<td>0.563</td>
<td>0.519</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldwin, Daugherty, Rowley &amp; Schwarz</td>
<td>1996</td>
<td>Medical school</td>
<td>916</td>
<td>1510</td>
<td>0.250</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bowers</td>
<td>1964</td>
<td>University</td>
<td>2,568</td>
<td>2,810</td>
<td>0.221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burns, Davis, Hoshino &amp; Miller</td>
<td>1998</td>
<td>University</td>
<td>77</td>
<td>151</td>
<td>0.045</td>
<td>0.077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>88</td>
<td>32</td>
<td>0.247</td>
<td>0.266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.332

0.192

0.05

0.221

0.045

0.077

0.247

0.266
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Level</th>
<th>N1</th>
<th>N2</th>
<th>r</th>
<th>r_s</th>
<th>z</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davis &amp; Ludvigson</td>
<td>1995</td>
<td>College</td>
<td>71</td>
<td>71</td>
<td>0.347</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis, Grover, Becker &amp; McGregor</td>
<td>1992</td>
<td>College</td>
<td></td>
<td></td>
<td>0.281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davis, Noble, Zak &amp; Dreyer</td>
<td>1994</td>
<td>College</td>
<td>39</td>
<td>10</td>
<td>0.147</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De Vries &amp; Azjen</td>
<td>1971</td>
<td>College</td>
<td>73</td>
<td>73</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erickson &amp; Smith</td>
<td>1974</td>
<td>College</td>
<td>68</td>
<td>50</td>
<td>0.335</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garfield, Cohen &amp; Roth</td>
<td>1967</td>
<td>College</td>
<td>50</td>
<td>30</td>
<td>-0.324</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huss, Curnyn, Roberts, Davis et al.</td>
<td>1993</td>
<td>College</td>
<td>142</td>
<td>78</td>
<td>0.088</td>
<td>0.194</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schab</td>
<td>1969</td>
<td>High school</td>
<td>580</td>
<td>835</td>
<td>0.055</td>
<td>0.124</td>
<td>0.464</td>
<td></td>
</tr>
<tr>
<td>Schab</td>
<td>1980</td>
<td>High school</td>
<td>580</td>
<td>520</td>
<td>0.15</td>
<td>0.052</td>
<td>0.204</td>
<td></td>
</tr>
<tr>
<td>Smith Ryan &amp; Diggins</td>
<td>1972</td>
<td>College</td>
<td>68</td>
<td>44</td>
<td>0.263</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stern &amp; Havlicek</td>
<td>1986</td>
<td>University</td>
<td>314</td>
<td>188</td>
<td>0.294</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thorpe, Pittenger &amp; Reed</td>
<td>1999</td>
<td>University</td>
<td>81</td>
<td>57</td>
<td>0.466</td>
<td>0.472</td>
<td>0.342</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>124</td>
<td>48</td>
<td>0.365</td>
<td>0.009</td>
<td>0.007</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>81</td>
<td>57</td>
<td>0.377</td>
<td>0.429</td>
<td>0.238</td>
<td>0.076</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>124</td>
<td>48</td>
<td>0.261</td>
<td>0.3</td>
<td>0.049</td>
<td>0.089</td>
</tr>
</tbody>
</table>
Criteria for including studies

Studies were included when they cited the proportion of males and females that cheated or they cited a statistic that could be converted to an effect size. Studies were excluded if they focused on the self-reported frequency of cheating, as this did not address the research question. As a result of this more stringent criterion, the reader should note significant variations in the studies included in this meta-analysis compared with that of Whitley, Nelson and Jones (1999).

A typical study using a single question was that of Smith, Ryan and Diggins (1972, p. 646). They asked “Have you ever cheated on an examination?” and 91% of men (N=44) and 97% of women (N=88) answered “Yes.” In other studies, more than one question was asked and the findings from each question were used in the meta-analysis. For instance, Antion and Michael (1983) used two separate questions answered yes or no from the Marlowe-Crowne Social Desirability Scale: (a) “I have never cheated on a test” and (b) “I have never used somebody else’s term paper.” They reported correlations between sex and these questions of –0.09 and –0.04, respectively, for 148 community college students. These results were used independently as indicators of cheating on tests and cheating on term papers in the meta-analysis.

In some studies the proportion cheating was determined for various situations (e.g., tests, assignments, plagiarism). In those cases where more than one proportion was cited, the highest proportion cheating in any one context was used as an indicator of the extent of cheating in a high school or university. For example, the study by Who's Who Among American High School Students (1994) provided details of the proportions of students who copied someone else’s homework (male = 63.2%; female = 72.2%), cheated on a quiz or test (male = 42.2%; female = 44.5%), or plagiarised part of an essay (male = 17.9%; female = 14.3%). The highest of the three proportions was used as an indicator of the extent of past cheating in high school. The separate questions were also used as indicators of cheating in particular contexts.

The study by Roth and McCabe (1995) used multiple questions and was typical of the reports not included in the meta-analysis. The dependent variable in their study was a composite measure. Students were asked how often they engaged in copying using crib notes, using unfair methods, or helping someone cheat. These behaviours were rated on a scale from 0 (never) to 5 (very often). The composite measure was the sum of the scale values. It focused on frequency of cheating and ratings and was included by Whitley, Nelson and Jones (1999) in their meta-analysis.

The study by Bonjean and McGee (1965) reported a larger percentage of males (0.716) than females (0.636) as actual or potential cheaters. It used a single measure but was not included. The report specified six cheating behaviours (seeking exam information from students, copying, collusion, lying about an absence, bringing information into an examination, purchasing a final exam). Students were asked “Have you or would you ever do this in the same situation?” Students answering “Yes” to any of the six situations were labeled as actual or potential violators. This study was excluded mainly because it contaminated past with potential future behaviour. Once again, Whitley, Nelson and Jones (1999) included this study in their meta-analysis.

Coding of studies

The studies that were selected were coded as follows (a) cohort (cheating in high school; or cheating in college and university); and (b) the specific context for cheating (e.g., cheating on tests or exams, plagiarism, copying assignments, lending work, and other forms of cheating). In this meta-analysis, gender differences in terms of the context were also examined.

The 21 studies included were published from 1964 (Bowers, 1964) through to 1999 (Thorpe, Pittenger & Reed, 1999). Two studies used community college students, 15 used college or university students, and four used high school students. The 14 studies listed in Table 1 cited the proportions of females and males who said they had cheated. In the meta-analysis (see Table 2) the studies varied in sample size from 49 to 206,865 and 56 out of the 64 effect sizes reported were from the USA. The total number of participants included for the meta-analysis were 108,358 females and 123,528 males.
Data analysis

The d statistic was used as the indicator of effect size. This is the standardised mean difference between men and women. Values of d are described as low (0.2), medium (0.5), or high (0.8) (Cohen, 1992). Differences in proportions were tested using the independent z-test and converted to chi-square with one-degree of freedom, then transformed to a correlation coefficient and an effect size. Studies such as Davis et al. (1992) reported only chi-square values for sub-samples and these were used to calculate a correlation and effect size. Studies that reported a probability such as p<0.01 were converted to two-tail z-scores and from there to chi-square, correlations and then effect sizes. In those instances where a study did not report a statistically significant gender difference, the effect size was categorised as zero following Whitley, Nelson and Jones (1999). There were varying numbers of participants in the studies analysed and the effect size was weighted by the size of each study. Where possible the results were checked against published findings; for example, the proportions in the Erickson and Smith (1974) article were used to recalculate the z-score quoted in the published paper (p. 109). A spreadsheet setting out the computations is available from the author upon request. The formulae for conversion of proportions into effect sizes are listed in Appendix A. In the following section, the results are discussed firstly in terms of the proportion of students who cheated and secondly in terms of the effect sizes for differences between males and females.

RESULTS

Proportion of male and female students cheating

The findings from earlier research confirmed that a large number of students had cheated. The overall proportions of female students cheating varied from a low of 0.05 to a high of 0.97 (median = 0.56) and for men, the proportion varied from 0.16 to 0.91 (median = 0.61). There was no statistical significant difference in the average proportions reported for males and females (t(34) = -0.58, ns). Accumulating the findings across the studies that reported both proportions and the actual number of males and females involved (N=226,003), showed that 21% of females and 26% of males had cheated. If the extremely large sample in the study by Astin et al. (1967) is excluded, then the proportions increase dramatically to 60% for both males and females. These studies showed a wide dispersion of findings (see Figure 1 for the distribution of proportions).

\[ \text{FIGURE 1: Distribution of the proportions of males and females cheating (N=18 studies).} \]

Effect sizes

The mean effect size between males and females for cheating in high schools was obtained from nine studies and was 0.14 (SE=0.03), and the mean effect size from 15 studies that considered cheating in college or university was slightly higher at 0.0.17 (SE=0.04). The overall effect size obtained from all studies (but using only the largest effect size where more than one was quoted) was 0.23 (SE=0.02). The effect sizes were not uniform but for the most part could be characterised as around 0.2 (see Table 3 for a listing of the average effect sizes for various assessment contexts).

\[ \text{TABLE 3: Mean effect sizes} \]

<table>
<thead>
<tr>
<th>Assessment contexts</th>
<th>N</th>
<th>Mean effect</th>
<th>H</th>
<th>chi-square</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The reader should note that the effect sizes for exams, quizzes, mid-term or unit tests were combined. This means that some studies provided more than one effect size and that the average effect size was not based on the ideal of independent observations. The average effect size for tests and exams was 0.186 and 0.141 if the results from the very large study by Astin et al. (1967) were excluded. This compared with 0.132 for the copying of assignments. The results from averaging the effect sizes from studies involving plagiarism involved eight effect sizes and had an average of 0.128. The remaining assessment context (lending work or other forms of cheating) involved eight effect sizes with an overall effect size of 0.315. Indeed, this was the largest effect size reported and points to an area that is worthy of further analysis. All of these effect sizes were plagued by problems of heterogeneity, and it is likely that a consistent average effect size has not been determined. The only exception appears to be that of the eight effect sizes cited for plagiarism, which appear to be relatively homogeneous and the net gender effect is very low.

The largest effect size of 0.332 for high school students came from a retrospective self-report cited by Baldwin et al. (1996) in their survey of medical school students. In fact, not one of the effect sizes for high school students came from a survey of that cohort. In contrast, the highest effect size for college students was 0.452 from a study of only 49 college students by Davis et al. (1994).

On exams, the largest effect size of 0.471 was reported by Baird (1980) from 200 college students; this study also reported the largest effect size (0.563) for copying. In fact, this was the largest of the 65 effect sizes calculated in Table 2. There were only two negative effect sizes; –0.242 reported for copying assignments by Antion and Michael (1983) and –0.324 reported for 50 female and 30 male college students by Garfield et al. (1967).

**DISCUSSION**

The findings from this review and evaluation of earlier studies indicated that substantial proportions of males and females engaged in cheating at high school and that substantial proportions continued cheating in college or university. It is a major concern for those involved in assessment that so many admitted cheating in some form or another. If we exclude the study of Astin et al., then the proportion of males or females admitting to cheating was 60% for both groups, reducing to 21% and 26% respectively for males and females if this study was included. Even these lower percentages (21% and 26%) must still be cause for some action.

The astute reader might note that in 12 of the 18 studies in Table 1 the proportion of males cheating outnumbered the proportion of females cheating. This is a case of Simpson’s Paradox, where inferences from large data sets are often the opposite of inferences from smaller sets. The paradox is caused by data from unequal sized groups being accumulated inappropriately into one large group and may also indicate the effect of some other intervening factor (e.g., sampling influences, methodology, disproportionate responding between males and females).
There was some support, however, for the view that there were small differences in the proportion of males and females cheating. The overall effect sizes for high school and college cohorts were low (d ≈ 0.2). The significant heterogeneity in the effect sizes means that our estimates are still sporadic, diverse, and have not yet yielded consistent findings.

These low effect sizes for male-female differences were consistent with the small effect size reported by Whitley, Nelson and Jones (1999). We used seven reports containing eight proportions that were not cited by Whitley, Nelson and Jones (1999) as well as ten studies with 30 effect sizes not contained in their article. Moreover, the basis of the classification of studies included in this paper varied from the approach outlined by Whitley, Nelson and Jones (1999).

Some limitations of this study arise from the fact that for the most part it focused on only two of the three forms of cheating identified by Cizek (1999), namely cheating by taking, giving, or receiving information from others and possibly cheating through the use of forbidden materials or information. It is not clear to what extent cheating by circumventing the process of assessment was covered. In addition, this paper has not covered observational and experimental approaches to cheating, and findings from these studies may validate or qualify the findings obtained.

A second limitation of this review is that cheating in contexts such as primary schooling, and adult and vocational education was not covered. Moreover, the effect sizes in the meta-analysis relied largely on the retrospective reports from college students on the extent of their high school cheating. These may not be accurate and may also be influenced by perceptions of what it means to cheat. In addition, almost all of the self-report studies were based on samples in the United States. Even a cursory knowledge of cheating practices worldwide indicates that the full dimensions of cheating across cultures might not be evident in the self-reports - mainly from the United States - that were reviewed in this paper.

The use of the confidential, anonymous, and private self-report surveys does have some advantages but it was also clear that females outnumbered male respondents in many of the studies. This occurred for 14 out of the 18 studies of proportions and for 20 out of the 25 groups with effect sizes. It may be that the propensity of females to answer such surveys on cheating is in some way linked with the extent of cheating among males or females. For instance a lower response rate may mask the prevalence of male cheating. The disproportionate participation of males and females is especially evident in the Who's Who (1993, 1994) studies of the attitudes of leaders in high schools. These studies had low response rates to a mail survey (3,177 out of 8,000 in 1994), and the final group comprised only 29% males. This limits seriously any conclusions that may be drawn about the extent of cheating. Finally, the findings from this review indicate the prevalence of cheating in a group, whereas an indication of the specific incidence may be more helpful for educators. For instance, Kerkvliett and Sigmund (1994) reported that only 1.9% of students admitted to cheating in a particular class.

This study confirms that cheating is a major educational problem and one that is likely to devalue assessment findings at all levels. Small differences between males and females were evident, but the effect of these differences was quite low. Both male and female students have cheated in large numbers, and unfortunately this affects many aspects of teaching, learning, and assessment and can disadvantage honest students.

Acknowledgments

An earlier version of this paper was presented at the International Conference on Measurement and Evaluation in Education, Universiti Sains Malaysia, Penang, November 2001. The helpful comments of two anonymous reviewers are gratefully acknowledged.

References


* references with an asterisk are used in the meta-analysis of results

**APPENDIX A**

**Formulae for conversion of proportions to effect sizes**

\[ Z\text{-score} \, z = \frac{(p_1 - p_2)}{\sqrt{(p^*q^*((1/n_1) + (1/n_2)))}} \]
Chi-square $c^2 = z^2$

Correlation $r = \sqrt{c / (n_1 + n_2)}$

Effect size $d = 2r / (\sqrt{1 - r^2})$

Descriptors: Meta analysis; Sex differences; Error, Cheating; Student Behavior; Student Evaluation; Academic Misconduct