

The Statistician as Collector

There is something about cricket that attracts collectors. Books, memorabilia, ephemera, photos, films and videos: all have enthusiastic adherents. It could be said that cricket statisticians have a similar mindset, except that we collect information. This information, while it can seem intangible, does originate in more physical things.

In my case, I collect copies of original Tests match scores. Early on in my studies of cricket, I realised that modern standards of great detail in records of Test matches were not matched by earlier historical records. If this could be recovered for many old matches, it would allow a more sophisticated assessment of historical changes in the game. Cricinfo began preserving ball-by-ball records of Tests in 1999. Earlier times were terra incognita: access was needed to original scores, as recorded by the scorers.

These scores, I found, are precious things. Hand-written, they are rarer than any published book, and most have never been copied in the past. Archiving them has been the work of a (very) few dedicated people at a few centres. The sad truth is that a remarkable number of original scores have been lost irretrievably: the many losses include (as far as I know) both the Tied Tests. Still, much survives, and I have been able to locate and copy scores of about one thousand of the 1,400 Tests played before 1999, and a significant number since. The originals remain with the archives; the copies have little intrinsic value, yet they preserve a unique record of cricket in earlier times.

The best archives of old scores are held at the M.C.C. Library at Lord's, and by Cricket NSW at the Sydney Cricket Ground. At both centres, researchers are welcome, and I would particularly like to commend Colin Clowes at Cricket NSW for his help over a number of years. By good fortune, Cricket NSW was bequeathed a large collection of scores by the legendary scorer Bill 'Fergie' Ferguson. Ferguson scored thousands of matches between 1905 and 1957, including many Tests not involving Australia. The players witnessed and recorded by Fergie range from WG Grace to Garfield Sobers, and he recorded every ball Don Bradman faced in Test matches. Evidently, when he was forced into retirement, much against his will, by the Australian Board of Control in 1953, he chose to send those scores still in his possession to the NSW authorities instead. This is probably fortunate, because Cricket Australia has evidently kept no scores older than 1980.

The earliest Ferguson score at Cricket NSW, a 1909 Ashes tour book, presents a puzzle. It is complete, but only as a photocopy. Who made the copy, and the whereabouts of the original, are not known. This does open the possibility that other old scores are out there. If any reader can shed light on this puzzle, or the whereabouts of any scores (including amateur scores), I would be delighted to hear about it.

The Melbourne Sports Museum holds several early Australian tour books, including Fergie's first book from the 1905 tour. There is also the earliest complete score for a Test in Australia, at Adelaide in 1884/85, found in the back of the 1884 England tour book. For many other tours, it comes as a surprise that many Tests are not represented by official scores. Much of the 'official' data for these Tests actually derives from press reports or *Wisden*. If tour scorebooks of the 1955 tour of West Indies and the 1956 nine-Test tour were kept, they do not appear to have made it back to Australia; they cannot be found today.

There are no central repositories of scores in most other countries. No central archives are known in South Africa, India, Sri Lanka, Pakistan or West Indies, and New Zealand has little material before 1982. Some of the gaps were filled by collections kept by individual grounds, or by individual collectors. John Kobylecky collected many England scores, and there was Bill Findall's inimitable collection of his own scores. In Pakistan, Shahzad Ali Khan has sourced many scores in Pakistan from the 1980s and 90s from the scorers for Pakistan Television. Charlie Wat collected score copies in the 1990s that otherwise might have been lost. Bernard Whimpress assisted in tracking down Adelaide scores from recent decades, and Allan Miller in Western Australia has kindly provided his collection. In Melbourne, Jack Cameron of the ABC kept his complete set of MCG Test scores since 1958 in his attic, and kindly copied them out for me. Appreciation of their value is not universally shared: when I visited Cameron, he turned to his wife, over tea and cupcakes, and said "See, I told you *someone* would be interested in that stuff one day."

Other important unique personal collections of scores and documents, such as those of B.J. Wakley, and roving international scorer Geoffrey Saulez, have been sadly discarded shortly after their deaths. Only the printed book collections were kept; valuable in their way, but containing nothing unique.

Occasionally, amateur sources come to light. In Adelaide, Graham Pellen kindly copied some of his scores from the 1970s, which had been otherwise lost from official archives. They were superbly written, and superior in quality to a great many professional scores that I have seen.

Some other archives are all but closed. Cricket Australia does have an impressive collection for Tests and One-Dayers since 1980; I gained a few hours access to this in 2013. Since then, the collection has been returned to deep storage, and negotiations to have this collection moved to Melbourne Cricket Club Library have not borne fruit.

Pooling together the sources, scores have been found for all Tests in Australia back to Adelaide 1973/74, and all Ashes Tests back to the First World War. Thanks to archives at each of the Test grounds, Tests in England are complete back to Headingley 1957. About ninety percent of all scores from the 1990s have been found. Unfortunately, some important gaps remain. Persistent enquiries have failed to unearth any full score for Sri Lanka's record six for 952 in 1997. It amazes me that no

one thought to preserve these original records at the time.

There are other curious gaps. The record for Australia's recent Tests is complete: rolling backward in time, scores or full ball-by-ball records for every one of Australia's Tests have been found, until we reach the Tied Test in Madras in 1986, which is lost. It is the only such lost Test since 1982; of all the ones to lose, this would have to be the most frustrating. There is a theory that an error was found in the score that proved the match was not a tie after all, and embarrassment was avoided when the score was conveniently 'lost'. There have been rumours (as opposed to hard evidence) of this for years: one Australian player told me he was convinced that Australia actually won the match by two runs.

If all this collecting is to have a point, the information must be made useable. Rendering the Test scores into digital computer records, ball by ball, is essential to extracting the greatest knowledge from this resource. This was the work of my spare time over a number of years, and now all known Test scores have been so treated, and combined into a one-ball-per-line database. When combined with the ball-by-ball records of Tests since 1999 at Cricinfo website, 82 per cent of Tests, comprising some 3.6 million balls, now have complete records.

Perhaps this is not the place for an exposition on what was found, but there are some interesting aspects that emerge from such detailed study. In particular, the number of apparent errors or anomalies in the surviving scores is surprising, occasionally disturbing. Problems seem to be more common looking back further in time, but scores from as recently as the late 1990s can present some serious puzzles. I would estimate that perhaps two-thirds of the Test scores that I have studied in detail contain some sort of anomaly.

There is a long list of anomaly types, including missing or extra overs, scoring strokes that don't add up to accepted totals (both batsmen and bowlers), mismatches between scoring strokes in batting and bowling sections (sometimes the numbers of ones, twos, threes and fours in the batting section do not match the bowling section), strokes possibly attributed to the wrong batsman. The devil is in the detail: it is actually rare to find differences between the gross totals in the scores on one hand, and the 'official' published scores on the other. An example of a more common problem is Charlie Macartney's legendary 345 versus Notts in 1921. The score gives Macartney 345, which is fine, but when re-scored into a ball-by-ball form, his score comes to 343. There is a 'rogue' shot for two that is not found in the bowling analysis.

Anomalies are more common when it comes to secondary statistics such as balls faced. There happen to be two surviving scores of a Test match at Hyderabad in 1984 (this is actually uncommon). Both show John Reid scoring 106, but the New Zealand score has him facing 325 balls, while the

Pakistan score shows 272 balls. That's quite a difference: the latter appears to be the accurate figure, although the former is the one found in online sources.

Generally these things are curiosities. Occasionally, there is potentially more substance to the problem. In the score of the final Test of 1928/29, when Don Bradman was batting with Jack Ryder, there is a hit for four attributed to Ryder that is in the wrong place. There are various possible ways to resolve this, but they include the possibility that the four was actually hit by Bradman. Readers will hardly need to be told what this would mean for Bradman's final batting average! In truth, the balance of probabilities favours the runs being hit by Ryder, but the tantalizing alternative remains: that the Don's real batting average was exactly 100.00.

There is a common source to many of these problems, in that many of these surviving scores are actually re-copies. Fergie himself spent his evenings copying out his linear sheets into traditional scores (sometimes exquisitely crafted). Without anyone to cross-check, errors crept in, but for many Tests, these copies are all that survive. Where both survive, the more prosaic and messy linear sheets are distinctly more reliable.

Other interesting anomalies include the matching innings of Barnes and Bradman, 234 each at the SCG in 1946/47. Barnes threw his wicket away shortly after Bradman, saying that he thought his innings would be remembered longer if his score was the same as the Don's; the incident has a place in cricketing folklore. Now, there are two surviving ball-by-ball records of this Test. In one, by S.G. Miller, Barnes' scoring strokes add up to 233, while in Fergie's score they come to 235. Does the folklore need revision? Fortunately, a full ball-by-ball re-score of the bowling figures restores the historical 234.

An intriguing case concerns the 1993 score at Adelaide Oval Tests, sensationally won by the West Indies by one run. Australia was out for 184 in the final innings: however, the batting strokes in the score add up to 185! Could it be that the match was really tied? On closer inspection it is not likely, as there is a stroke for two (in the bowling section) that seems to have been mistakenly recorded as three in the batting. The two runs was confirmed by Graham Pellen from his own score of the match, and that version has the virtue of keeping the batsmen at the right ends.

(The score helps reveal one other oddity about this match. For most of Test history, no balls only incurred a penalty run if no other run was scored. In 1998, in a small victory for logic, the protocol was changed, to add the extra run in all circumstances. A no ball hit to the boundary now incurs five runs, where once there were four. Had this protocol been in place in 1992/93, Australia would have won that Adelaide Test, because the Australians scored off seven no balls to the West Indies four.)

The lesson from this and many other examples is that in most cases, the best resolution of anomalies, when they arise, is acceptance of the status quo. Still, feedback from other statisticians suggests that some find this all a bit disquieting. The idea that there might be 'error bars' around the career stats of past players doesn't always sit well. Did Victor Trumper actual score 187 not out at the SCG in 1903/04, as the surviving score suggests, rather than 185 not out? Perhaps we should be happy to live with this level of uncertainty, but not everybody is comfortable with this. It certainly would be reassuring to know that Wally Hammond really scored 336* at the Basin Reserve in 1933, rather than the 337* suggested by a close study of the score.

In truth, the sources for Test match scores of long ago are not always unimpeachable. In general, however, the uncertainties are not profound. Whatever the problems, collecting as much information as we can about these matches can only enhance our understanding and appreciation of the game's history.

The author would be delighted to hear from anyone who may know of full scores, even unofficial one done by amateurs, of old Tests matches, particularly before 1980. Contact: stats334@iprimus.com.au

Charles Davis

April 2014