ILAG 2001

Papers

International Legal Aid Group

Avrom Sherr

Peer Review and
Model Clients:
The English Experience

Melbourne
Australia
13-16 July

PEER REVIEW AND MODEL CLIENTS: THE ENGLISH EXPERIENCE

Avrom Sherr

Peer review of Contracted Work (Chapter 5)

This chapter is the first of several assessing the quality of work conducted under the contracting pilot. Later chapters discuss results from the client survey, model clients, contract outcomes and a management review. This chapter focuses on the results from the peer review of contracted work.

Peer review of contracted work on cases

Peer review took place on 718 contracted cases. The review data was matched with BriefCase data for those cases and data on groups. Reviews took place in five work categories: debt, employment, housing, personal injury and welfare benefit cases. Fifty-two contractees were reviewed. In this report, this data set is referred to as the *main peer review data set*.

During each peer review, a sample of the main peer review data set was second-marked to provide a cross-check on the peer review scoring. A total of 173 files were double marked in this way. This is referred to as the *double-marked data set*. Additionally eighteen sets of model client reports were also marked by five peer reviewers (the peer reviewed model client data). They are discussed here in relation to assessing the reliability of peer review, although the substance of these results is discussed in Chapter 7 on model clients.

This chapter first discusses in detail the reliability and consistency of peer review (paras 5.5 – 5.33). It then discusses peer reviewers' assessment of quality under the pilot.

Reliability of peer review as an indicator of quality

It should be emphasised from the outset that this detailed an analysis of peer review marking has not to our knowledge been attempted before in a socio-legal context. Observer reliability, for instance, is usually assumed or deduced from very generalised analysis. The reliability of peer review has been criticised,[1] whilst also being regarded, largely on the basis of intuition, as the 'gold standard' of quality assessment.[2] It is also thought to be capable of

---

[1] See Daniel, H.D., (1993), *Guardians of Science: Fairness and Reliability of Peer Review*, (VCH Verlagsgesellschaft MBH, Weinheim). See, Sherr *et al* (1994), *The Quality Agenda: Volume I* , p. (HMSO, London).

[2] Canadian Bar Association (1987), *Legal Aid Delivery Models*, Canadian Bar Association Standing Committee, National Legal Aid Liaison Committee, pp.91-92

providing insights into expertise which is not a feature of other forms of quality tool.[3] Given these competing assessments of peer review, it was important to examine the consistency and reliability of peer review as a measure of quality.[4]

As outlined above, 173 cases were marked twice (i.e. by two peer reviewers). The double-marked data set contains two sets of marks for each case. The 'original' mark was given for a case by a peer reviewer after considering the file.[5] The second assessment was conducted by another reviewer who was aware that the file had already been marked by another peer reviewer but was not aware of what that mark actually was. A limitation of this approach was that the double-marker would have been conscious that they were second guessing the original review and may have been more cautious as a result. Nevertheless, this provided the most practicable large-scale check on peer reviewer consistency.[6]

The purest way of testing inter-marker reliability was the peer reviewed model client data. Peer reviewers were asked to assess eighteen sets of correspondence and model client reports produced under the model client aspect of the research. As a result, eighteen model client reports were assessed five times, once by each peer reviewer. This is the equivalent of ninety reviews. The advantage of this review was that it allowed a comparison of peer reviewer approach based on identical information and in circumstances where no reviewer was consciously 'second-guessing' another reviewer.

Reliability of the instrument

The reliability of the peer review instrument (the checklist) was examined to see if reliability was a particular issue for certain questions on the peer review check-list. Most questions indicated satisfactory levels of agreement between the peer review markers, but some did not. This provides some interesting

---

[3] This is because expert knowledge is thought to take three forms: declarative knowledge (the stating of concepts and categories in a propositional form), procedural knowledge (the ability to recall and use declarative knowledge) and expertise (where the abstractions of declarative and procedural knowledge are superseded by the ability to perceive and act on problems so that, "a so-called 'given' problem is not really given since *it is seen differently* by an expert"). See, Estes in Lesgold and Glaser, *Foundations for a Psychology of Education* (1989, Hillsdale, New Jersey), p. 6 *et seq*. See, also, de Groot in Bransford *et al* in Lesgold and Glaser, p.203. Schõn refers to tacit knowledge or "knowing in action", but suggests that although, through tacit knowledge, experts can recognise good or bad work there may be greater difficulty in articulating in a 'technical-rational' form *why* such work was good or bad. Schõn (1983), pp.49-69.

[4] See Daniel, H.D., (1993), *Guardians of Science: Fairness and Reliability of Peer Review*, (VCH Verlagsgesellschaft MBH, Weinheim).

[5] They would have been aware of the possibility that any of their files could be double-marked but were not aware of which files would be double-marked.

[6] There was no practical alternative to this that could fit with the other aspects of peer review and represent a reasonable burden on contractees.

insights into what may (and what may not) be susceptible to peer review from contractees' advice and assistance files.

To assess the reliability of the instrument, the original and second mark was compared *for each question on each file*.

All questions answerable using a five-point scale (grades 1 (poor) to 5 (excellent)) were compared using Spearman's rank correlation coefficient. This shows whether or not peer reviewers were broadly marking files in the same direction, (i.e. tending to higher or lower marks on the same files).[7] Statistically significant correlation coefficients were found for all of the peer review checklist criteria save the question, "How effective in achieving clients' wants/needs through negotiations [was the advisor]?"

Kappa statistics were calculated for questions which were answerable 'yes' or 'no' as it is not appropriate to use correlations on such data. As an additional test on the five-point scale questions, Kappa statistics were also calculated to provide an additional indication of the extent to which the peer reviewers agreed with each other in relation to specific marks on specific files on the questions marked on a five-point scale.[8]

Kappa coefficients on the five-point scales generally indicated significant Kappa coefficients which were also positive, however the Kappa coefficient was generally quite low. As a result, the results were re-coded to collapse the five-point scale into a three-point scale. Kappa coefficients were then calculated for each of the individual criteria marked by the peer reviewers on a three-point scale. In general, these also indicated significant Kappa coefficients and the Kappa coefficient itself was improved (indicating greater reliability where the scale is reduced from a five-point scale to a three-point scale when interpreting the results).[9]

The Correlation and Kappa coefficients are summarised in Table T5.1. On the Kappa coefficients, some of the criteria did not show significant agreement and/or a positive kappa coefficient. As a result there was no significant agreement between original and second-markers on certain questions. This suggests that it is difficult to assess such criteria by peer review on advice and assistance files. This applied to the following questions:

Has the adviser understood the position?

---

[7] Cramer, D. *Fundamental Statistics for Social Research*, Routledge, London (1998) suggests using Pearson's correlation coefficient but as the data we are using is non-parametric, Spearman has been used. Cramer (1998) also suggests that Kappa tests are best used on categorical data where there are two judges to compare.

[8] Kappa also takes into account the pattern of results. For there to be agreement, the Kappa coefficient (K in the tables below) needs to be positive and also statistically significant. Ideally, the Kappa coefficient should be in excess of 0.7 for agreement to be marked. For example, the coefficient improved from 0.13 to 0.24 for the overall marks on each file.

[9] For example, the co-efficient improved from 0.13 to 0.24 for the overall marks on each file.

Was further fact-finding appropriate?

Was further fact-finding efficiently executed?

How effective [was the contractee] in achieving clients' wants/needs through negotiations?

Some questions had positive Kappa coefficients and indicated trends ($p < 0.10$) rather than statistical significance ($p < 0.05$). Generally, where the level of agreement was only indicative of a trend, the questions applied to a small number of cases only. The small sample size would have had the effect of reducing the significance and therefore rendering testing of the level of agreement more uncertain. This applied to these questions as a result:

If no other work, was this appropriate?

If no disbursements, was this appropriate?

Referral considered/advised/acted on?

The problematic questions are included in the report for completeness of analysis but warnings are given at the appropriate points. All other questions have been included as being reliable indicators of quality when marked by peer review.

Consistency in rating individual files

As well as assessing consistency on individual questions, it is important to examine the consistency between peer reviewers in their overall assessment of files.

The double-marked data set was used a means of investigating in more detail consistency between peer reviewers. The next table provides an indication of the extent to which the original scores of each peer reviewer agreed with the second-marker's scores for the same file. The following table summarises the overall position.

Table 5.1: Crosstabulation of original and second marks

	Overall (original mark)						
Overall (second-mark)	Lowest	2	3	4	Highest		Total
		n			n		
Lowest	11%	2	6%	0%	0%	0%	3
2	67%	8	22%	14%	5.0%	13%	28
3	11%	23	64%	56%	35.0%	38%	86
4	11%	3	8%	29%	48%	50%	50
Highest	0%	0	0%	1%	13%	0%	6
N	9		36	80		8	173
Percentage within 1 mark	78%		92%	99%	95%	50%	

Tests: Spearman Correlation .443, p<= .0001, Kappa .134, p<= .005.

This table highlights areas of agreement and disagreement between peer reviewers. At the middle range of the five-point scale agreement within one mark was extremely high (92% to 99%). At the extremes of the scale, agreement was much less marked, particularly when identifying higher quality work. There was also a notable tendency, especially amongst second markers, to use the middle of the marking range.

There is a highly significant correlation between the marks of the first and second markers, and a highly significant Kappa coefficient, although the value of the coefficient itself is low. The Kappa scores increase when the five-point scale is collapsed to a three-point scale.[10]

Although the levels of agreement are quite high, because of low Kappa scores there are still some concerns about the ability of peer review to produce consistent indications of the quality of individual pieces of work. This problem is investigated in more detail in the following table which summarises the scores for each peer reviewer and the equivalent 'second marks' on their files.

[10] Spearman Correlation, .402, p <=.0001; Kappa, .337, p <= .0001

Table 5.2: Peer reviewers' scores and the relevant second-marks for the files they marked (double-marked data set only)

		Overall Mark			
		Below Standard	Standard	Above Standard	N
PR1	Original Mark	26%	52%	22%	23
	Second mark	17%	48%	35%	23
PR2	Original Mark	22%	31%	47%	36
	Second mark	11%	39%	50%	36
PR3	Original Mark	62%	18%	21%	34
	Second mark	27%	44%	29%	34
PR4	Original Mark	12%	78%	10%	41
	Second mark	20%	61%	20%	41
PR5	Original Mark		78%	22%	18
	Second mark	11%	61%	28%	18
PR6	Original Mark	24%	24%	52%	21
	Second mark	19%	48%	33%	21
All Cases	Original Mark	26%	46%	28%	173
	Second mark	17%	51%	32%	189

The table shows two things. Firstly, it indicates a level of difference between the original overall scores on a file and the equivalent scores by the second-marker. PR3's results are worthy of particular attention. This reviewer felt that 62% of the files they marked were below average standard. This was noticeably above average for all reviewers (26%, even including PR3's scores). Equally, the second-marking of PR3's files tends to support the view that those files were of a poorer quality than most other files. Nevertheless, the level of difference between PR3's original scores and the second-markers for PR3's files is cause for concern. PR6 by comparison marked more generously than the second-markers felt was merited.

Secondly, the 'All Cases' row suggests some interesting differences between peer reviewer behaviour when marking files for the first time and when they are second-marking a file. As noted above, when second-marking, on the whole peer reviewers appeared more likely to give files the benefit of the

doubt and to indicate files were at or above a standard level of quality than the first-markers of the file.

This seems to suggest that some disagreement between peer reviewers was caused by peer reviewers approaching the second-marking task differently from the way that they approached the original marking of files. They may have done this because they had less time, it was the end of a day of marking files and they were aware that they were 'only' cross-marking these files. Or they may have done so out of a concern not to risk judging a file differently from another peer reviewer.

Because of this difference in approach, the double-marked data set may exaggerate the difference between peer reviewers and suggest that the Kappa coefficients underestimate the extent to which peer review provides a predictable and 'accurate' assessment of quality on individual files. This possibility was tested using the peer reviewed model client data set.

Model client files assessed by peer reviewers

Peer reviewers were asked to assess model client reports (see Chapter 7) on a scale of 1 to 5 (1 being poor and 5 being good). Eighteen files were reviewed by five of the peer reviewers. The following table provides Spearman's Rank Correlation and Kappa scores comparing each peer reviewer's assessment of the model client files.[11]

Table 5.3: A comparison of peer reviewer scores of model client documents

Peer reviewers	Spearman's rho coefficient	Sig.	Kappa coefficient	Approx. Sig.
1 and 2	.812	.0001	.325	.008
1 and 3	.464	.053	.327	.040
2 and 3	.588	.010	.270	.052
1 and 4	.651	.003	.443	.007
2 and 4	.636	.000	.302	.002

[11] The Kappa score was on the basis of a three-point scale.

3 and 4	.574	5.013	.280	.068
1 and 5	.557	3.016	.589	.0001
2 and 5	.718	6.001	.493	.001
3 and 5	.684	1.002	.486	.004
4 and 5	.785	2.0001	.571	.0001

Significant Spearman's Rank correlation coefficient indicated whether or not peer reviewers are broadly marking files in the same direction, (i.e. tending to higher or lower marks on the same files). The closer the coefficient was to 1.00, the closer the reviewers were to total agreement. Statistically significant correlation coefficients for each pair of peer reviewers were apparent save for the pairing of Peer reviewer 1 and 3. The fact that the coefficients were also high, indicated that the level of agreement was high. Similarly, the Kappa coefficients were calculated on a three-point scale. On these marks the Kappa scales were all positive, most were statistically significant and were also generally fairly strong. They were notably higher than the scores from the main double-marking exercise.

As a result, where the peer reviewers were looking at the same documentation on the same case and without being conscious that they were conducting a "cross-check" of other peer reviewer's work, the peer reviewers exhibited more significant levels of consistency. These levels of consistency exceeded those found in the double-marked data set.

There were some limits to this finding however. Peer reviewer 3 was not as likely to agree with other peer reviewers. Similarly, differences in peer reviewer scores on individual files are sufficient for a recommendation that if using peer review more extensively, the Commission should not rely on one peer reviewer's assessment of individual files or even a small number of files.

The next stage of this assessment was to look at peer review judgements averaged out over a number of files of peer review reliability.

Consistency in rating contractees

It was also possible to examine how peer reviewers rated contractees by looking at the reviewers' (average) mean score for each contractee, rather than individual files. This helped to even out some of the effects of looking only at one file. Table T5.2 summarises the position.

For 11 contractees out of 53, the reviewers' average scores were one mark or more apart. Otherwise the marks were generally close to each other, even though only a very small number of files (usually 2 or 3) had been double-marked for each contractee (hence there was only a limited opportunity to 'even out' differences between each reviewers scores). Where the sample size (n) looked at was larger consistency was even more marked. The correlation coefficient suggested strong levels of agreement.[12]

Conclusion on Peer review Methodology

Peer review scoring has been rigorously investigated. The instrument itself was checked to ensure that there was reliability for each question on the checklist. The majority of questions produced satisfactory reliability between markers. Some questions have been identified as needing to be treated with caution. This caution is emphasised at relevant points below.

Overall, the assessment of peer reviewer reliability suggested that reviewers tended to agree with each other in their general assessments of quality, but that such agreement was only sufficiently robust when looking at assessments of particular contractees across a number of files rather than individual cases. Whilst the results indicate the care that needs to be taken in setting up peer review (and structuring peer review criteria), providing training and monitoring and so on, for the purposes of this research the results have been strong, and even in terms of assessing actual contracting decisions, peer review could be developed on the basis of ensuring peer reviews were handled by more than one peer reviewer who assessed firms on the basis of a number of cases.

Overview of results from peer review

Peer reviewers were asked to answer a number of questions yes or no and the remaining questions on a scale of 1 to 5. Table 5.4 (below) summarises the results of the yes/no questions. The 'No' column indicates the level of quality concern where the question applies. Some questions were only applicable to a smaller number of cases. To put the figures into the perspective of all peer reviewed cases, an adjusted figure is presented in the column "No (% of all cases)" to give an indication of how serious the problem is when considered against the total number of contracted cases. As an example, in 190 cases peer reviewers found that there was a need to deal with referral. In 64% of these cases there was a failure to deal with referral properly. Because a referral

[12] Spearman's Rank Correlation .579, p<=.0001. It is not possible to produce a Kappa score for scale-data (i.e. integers which can occur at any point on a scale) rather than ordinal data (scaled 1, 2 or 3).

issue does not arise in every case, it might be argued that this figure overestimates the seriousness of the issue.[13] To take account of this, the failure rate as a proportion of all contracted cases was calculated. This adjusted figure is 17%. Thus in 17% of all cases an adviser failed to consider, advise on, or act on an effective referral.

Table 5.4: Peer review results (general summary 1: yes no questions)

	Yes (%)	No (%)	No (% of all cases)	Total Valid Responses
Does adviser appear to have understood the problem ?*	96	4	4	713
Was the advice given in time/at the right time?	92	8	8	689
If no other work was carried out, was this appropriate?*	60	40	11	194
Were any disbursements incurred appropriate?	91	10	1	63
Were any disbursements incurred necessary?	81	19	2	63
If no disbursements were incurred, was this appropriate?*	94	7	5	604
Did the adviser consider, advise on, act on an effective referral to other organisations?*	36	64	17	190

This was one of the criteria where peer reviewers did not show significant agreement in their approach to assessing this issue. The results on this question must therefore be treated with caution.

Broadly speaking these results are encouraging, most advisers appeared to understand the problem (96%)[14] and to give timely advice (92%).

Disbursements, where incurred, were usually appropriate (91%), but were less likely to be regarded as necessary by the peer reviewers (in 19% of cases where disbursements were incurred the disbursements were regarded as unnecessary, although this only amounted to 2% of all peer reviewed cases). In 7% of cases where disbursements were not incurred, the peer reviewer felt disbursements ought to have been incurred. When looked at as a proportion of all contracted cases (as seen by peer review), this suggests that in 1 in 20 of all cases contractees had failed to incur necessary disbursements. There are no benchmark figures to help assess whether this is a problem that occurred prior to contracting or as a result of contracting, but these figures suggest that, in the peer reviewers' judgement, disbursements were being under- rather than over-incurred.

[13] This report does not argue in favour of, or against that view, but presents the results in both ways to get a sense of the level and depth of each quality concern.

[14] Variability in peer reviewer approach to this criterion suggests this figure needs to be treated with caution.

Of more concern was the finding that in 40% of cases where no further work was done on a file (beyond the first interview), peer reviewers considered this to be inappropriate.[15] This amounted to 11% of all cases (or over one case in 10).

Similarly, it was worrying to see that in 64% of cases (or 190) peer reviewers felt that the contractee should have considered, advised, or acted to ensure an effective referral but did not do so.[16] Therefore, in 17% of cases reviewed by the peer reviewers (almost 1 in 6 of all contracted cases) referral was not dealt with adequately by contractees.

Since this pilot, the Quality Mark and (in some parts of the country) Community Legal Service Partnerships have been introduced which has placed additional emphasis on referral.[17] Work on referrals has not, to date, provided any convincing indication of a benchmark level for referrals. This research suggests that referral should be considered and/or acted on in about 26% of cases (1 in 4 cases); and that there has been a failure to act appropriately in the majority of those cases (1 in 6 cases). This might provide a benchmark indication of referral activity to be expected by CLSPs.[18]

The next table looks at the other criteria by which peer reviewers assessed cases. The shaded columns provide summary figures for the percentage of cases that were marked below and above threshold competence.

---

[15] Variability in peer reviewer approach to this criterion suggests this figure needs to be treated with caution.

[16] Variability in peer reviewer approach to this criterion suggests this figure needs to be treated with caution.

[17] See *Legal Aid Franchise Quality Standard*, Legal Services Commission, Fourth Edition April 2000 p.46 at L2.5 Limits of Professional Competence and Referral. See Moorhead R., *The Community Legal Service Pioneers in Practice Research Report*, Lord Chancellors Department, March 2000, chapter 7, pp 117-145.

[18] But see Moorhead R., *The Community Legal Service Pioneers in Practice Research Report*, Lord Chancellors Department, March 2000, chapter 7, pp 117-145.for a discussion of the difficulties of applying such indicators to individual organisations.

Table 5.5: Peer review results (General summary 2: five-point scale questions)

	Non-performance		Total -ve (%)	Threshold competence	Total +ve (%)		Excel-lence	
	1	2	-	3	+	4	5	N
How effective were the adviser's communication and client handling skills?	2	10	12	54	34	29	5	717
How effective were the adviser's fact and information gathering skills?	2	15	17	46	38	30	8	717
How legally correct was the advice given?	3	19	22	47	31	24	7	714
How appropriate was the advice to the client's instructions?	2	18	20	48	31	24	7	714
How comprehensive was the advice?	4	29	32	40	28	22	6	716
Was further fact-finding work carried out appropriate?*	4	12	16	53	31	24	7	531

Percentage of all cases			*12*					*718*
Was further fact-finding carried out efficiently executed*	5	14	19	48	33	26	7	5311
Percentage of all cases			*13.7*					*718*
Was any other work carried out appropriate	4	10	14	59	28	24	5	618
Percentage of all cases			*13.8*					*718*
Was other work efficiently executed?	4	12	16	54	30	25	5	619
Percentage of all cases			*13.8*					*718*
How effective in achieving what the client reasonably wanted/needed; was any work carried out through: a) letter writing and forms?	6	16	21	48	31	25	6	677
Percentage of all cases			*19.9*					*718*
b) telephone calls?	13	8	21	59	20	16	3	447
Percentage of all cases			*13.2*					*718*
c) negotiations?*	18	13	31	51	18	15	3	417
Percentage of all cases			*12.5*					
How effectively was the client informed of merits (or not) of claim?	4	25	29	45	26	21	5	715
How effectively was client informed of all developments	7	16	22	54	24	20	4	611
Percentage of all cases			*19*					*718*
Throughout the file did the								

organisation make an effective use of resources?	3	19	22	48	31	25	5	717
Overall Mark	**3**	**22**	**25**	**47**	**28**	**23**	**5**	**718**

This was one of the criteria where peer reviewers did not show significant agreement in their approach to assessment. The results on this question must therefore be treated with caution.

In general, these results are positive. Overall, advice was rated at or above threshold competence in 75% of cases looked at by the peer reviewers. The majority of this was rated as being threshold competence (47%), but over 1 in 4 cases were rated at competence-plus or excellent quality (28%).

On average, the peer reviewers thought that contractees performed well on communication and on the effectiveness and appropriateness of their fact-finding.[19] Contractees performed less well on the comprehensiveness of the advice (32% of advice was rated poorly compared with 28% being rated as good). Effectiveness in keeping the client informed of the merits was also rated at below competence (29%) more often than it was rated highly (26%). Advice was rated poorly for legal correctness in 22% of cases (over 1 in 5 cases). Effective use of resources was rated poorly in 22% of cases, but was rated positively in 31% of cases.

Differences between solicitors and the NFP sector

Table T5.1 summarises the results comparing the private practice sector with the NFP sector. As the distributions of the overall scores on each file shows, NFP agencies were rated more highly than solicitor agencies, as Table 5.6 and the accompanying bar chart shows.

Table 5.6: NFP, solicitors and groups overall peer review score compared

Overall Score	1 Poor	2 IPS	Below Threshold Competence	Threshold Competence	Above threshold Competence	4 Competence Plus	5 Excellent	N
NFPs	1%	24%	25%	34%	42%	28%	13%	188
Sols	4%	21%	25%	52%	23%	21%	3%	530

[19] The results for appropriateness of fact-finding need to be treated with caution because peer reviewers did not show significant agreement in their approach to assessing this issue.

All Cases	**3%**	**22%**	**25%**	**47%**	**28%**	**23%**	**5%**	**718**

Figure 1: Peer Review Scores NFPs and Solicitors Compared

The proportion of poorly-rated cases was the same but the difference between above threshold competence cases (23% for solicitors compared with 42% for NFP agencies) was highly significant (p= .0001). In simpler terms, this suggests that, in peer reviewers' assessments, about 1 in 5 clients got a higher level of service when they went to an NFP agency compared with those that went to a solicitor's firm. Conversely, the chances of getting substandard service were similar for NFPs and solicitors. This general phenomena held for most of the detailed peer review criteria. Statistically significant differences consistently indicated that the peer reviewers regarded the NFP agencies as providing a higher level of service (see Table T5.1). The one exception to this was the NFPs performance at negotiation.[20]

Differences between groups

Tables T5.1 and T5.2, also contain statistics comparing in detail the performance of each of the four payment groups (Groups 1 to 3 and the NFP sector). This section of the report provides a summary analysis of those statistics focusing in particular on statistically significant results.

Group 1 compared to the other groups

Peer reviewers' assessments of Group 1's work differed statistically significantly in the following respects.

Compared with Groups 2 and 3, Group 1 firms were significantly better at giving advice at the right time (which they did in 96% of cases) than Group 2 (90%)[21] and Group 3 (85%).[22]

Disbursements incurred by Group 1 were significantly more likely to be regarded as necessary by the peer reviewers for Groups 2 and 3.[23] Group 1

---

[20] This was one of the few questions where there were significant concerns about the reliability of the peer review assessment because this was an area where there were no significant levels of agreement between peer reviewers in their assessments of files.

[21] Chi-square p = 0.026.

[22] Chi-square p = 0.001.

was also significantly less likely to fail to incur disbursements where disbursements were thought to be appropriate by the peer reviewers. [24]

The overall picture shows that Group 1 was assessed at about the same level of quality as the other solicitor groups, with some slight improvement on Group 3. A comparison with Group 2 is more complex. Whilst scoring better than Group 2 on some individual criteria, Group 1 tended to get more below threshold competence scores than Group 2 but also more above threshold competence scores than Group 2.

The differences between Group 1's performance and the performance of the NFP agencies was more marked. Group 1 firms were significantly poorer than the NFP agencies in terms of advice giving: the legal correctness of the advice;[25] the appropriateness of the advice to the client's instructions;[26] and the comprehensiveness of the advice were all rated significantly more poorly for Group 1 cases than for NFP cases.[27]

Similarly, the carrying out of further work was generally rated more poorly on Group 1 cases compared with NFP agencies. Further fact-finding work was rated more poorly;[28] and Group 1 contractees were significantly less likely to be regarded as good at achieving the clients wants/needs through letter writing and form-filling than the NFP agencies.[29] Similarly, Group 1 was significantly poorer than NFPs at informing the client of the merits (or not) of the claim;[30] and was not as good at informing the client of developments.[31]

---

[23] Chi-square p = 0.035. All disbursements incurred in Group 1's peer reviewed cases were regarded as necessary. For Group 2 71% and Group 3 73% of their peer reviewed disbursements were regarded as necessary. Disbursements were incurred only in a relatively small number of cases.

[24] In 3% of applicable cases Group 1 failed to do this, compared with 13% of applicable cases for Group 3), p = 0.003. However, the number of cases where peer reviewers regarded this as applicable is small.

[25] Chi-square p = 0.021

[26] Chi-square p = 0.030.

[27] Chi-square p = 0.038.

[28] 21% of cases were negatively rated, compared with 11% for NFPs and 29.3% of cases were positively rated compared with 46% for NFPs, p = .003.

[29] Both groups had similar negative ratings (of about 20%), but the NFPs got more positive ratings, 47% rated positively compared with 26% for Group 1, p = 0.005.

[30] 33% negative rating, compared with 23% for NFPs, 19% positive rating compared with 27% for NFPs, p = 0.001.

[31] Negative scores were similar (c. 20%) but NFPs scored positively in 44% of cases compared with 21% of cases for Group 1 cases, p = 0.005.

Group 1 was significantly less likely to consider, advise, and/or act on an effective referral to other organizations[32] and was also rated significantly more poorly on effective use of resources when compared with NFP agencies.[33]

As well as these significant differences, the differences on most of the other criteria, whilst not statistically significant, support the view that the NFPs in the pilot, on the whole, performed to higher standards than Group 1 firms. The NFP agencies consistently outscored Group 1 in about 15% of cases,[34] which roughly translates to about 1 in 7 NFP clients getting higher levels of service. However, if one were to look purely at differences in the number of cases which were assessed negatively (below threshold competence) then there would be little difference between Group 1 and NFPs. As a result, the difference in quality perceived by peer reviewers is attributable to more numerous examples of service above threshold competence in the NFP sector.

Group 2 compared to the other groups

Group 2 differed statistically significantly from other solicitor groups in the following respects. There were two criteria on which Group 2 performed significantly better than Group 3, although both these results need to be treated with some caution.[35] Failure to carry out other work was felt to be inappropriate in far more cases for Group 3 (54% of applicable cases), compared with Group 2 (21% of cases) and failure to incur disbursements which the peer reviewers regarded as appropriate occurred in 13% of cases in Group 3, but only 3% of cases in Group 2. The differences between Group 2 and Group 1 are discussed above (Group 1 scored more highly on some criteria than Group 2 but also had more cases rated below threshold competence overall).

The NFP agencies performed significantly better than Group 2 on most of the relevant criteria. The adviser's communication and client-handling skills at the interview were rated significantly better in the NFP sector;[36] as was the

---

[32] Peer reviewers assessed Group 1 as not doing this in 74% of applicable cases, compared with 35% of NFP cases, p = 0.001. Peer reviewer variability in marking this criteria means the results should be treated with caution.

[33] NFPs scored above threshold competence marks in 42% of cases compared with 24% for Group 1, p = 0.023. Negative scores were similar.

[34] This was true of all criteria save achieving the client's needs/wants through negotiation where the NFP sector was outscored in about 18% of cases, although peer reviewers tended to disagree in their assessment of organisations on this criterion and so results have to be treated with extreme caution here (See Appendix B).

[35] These were both criteria where peer reviewer variability means that these results should be treated with caution.

[36] Communication and client-handling skills (poor ratings were similar at about 10%, but positive ratings were lower for Group 2 (23%) than for NFP agencies (44%), p = 0.002).

appropriateness of advice to the client's instructions.[37] The appropriateness and efficient execution of further fact-finding and other work were all rated significantly more highly in the NFP sector;[38] and, the effectiveness with which clients were informed of the merits (or not) of a claim (poor ratings were similar at about 23%, but positive ratings were lower for Group 2 (18%) than for NFP agencies (38%)).

In a number of these areas however Group 2 had fewer negatively assessed cases than the NFPs but the effect of positive ratings was significantly stronger for the NFPs and so Group 2 as a whole did more poorly than the NFP sector in these areas: the appropriateness of other work;[39] the efficiency with which further work was done;[40] and, the effectiveness with which clients were informed of developments.[41]

Conversely, in cases where no further work was carried out beyond the first interview this was felt to be inappropriate in 53% of NFP cases, compared with 21% of Group 2 cases. This difference was also statistically significant.[42]

Group 2 performed significantly more poorly than NFPs on considering, advising and/or acting on effective referral.[43]

Looking beyond these significant results, there are some other factors which should also be emphasised. Group 2 consistently had the lowest numbers of cases marked at below threshold competence (even when compared with the NFP sector). Under peer review, a view of quality that only measured 'threshold-competence' and did not encourage or evaluate higher levels would indicate that Group 2 was the best performing group under the pilot.

In general, Group 2 performed similarly to Group 1 and slightly better than Group 3, with an average of 14% of its cases scoring more poorly than NFP

---

[37] Poor ratings were similar at about 18%, but positive ratings were lower for Group 2 (22%) than for NFP agencies (42%), p = 0.050.

[38] Chi-square p = .00011, 0.013, 0.002 and 0.004 respectively. See Table T5.2 for detailed percentage breakdowns.

[39] 7% of Group 2 cases were rated poor compared with 12% for NFP agencies, but the effect of positive ratings was stronger with Group 2 having 15% rated positively compared with 38% for NFP agencies

[40] Group 2 had 9% compared with 15% for NFPs, but the effect of positive ratings was still significantly stronger with Group 2 having 19% rated positively compared with 43% for NFP agencies.

[41] Group 2 had 13% of cases rated negatively compared with 23% for NFPs, but the effect of positive ratings was still significantly stronger with Group 2 having only 5% of cases rated positively compared with 44% for NFP agencies.

[42] Chi-square p = 0.002.

[43] They failed to do this in 79% of applicable cases compared with 35% of NFP cases, chi-square p = .00011. This is a criterion where peer reviewer variability means the results should be treated with caution.

cases on each criteria (compared with 15% for Group 1 and 19% for Group 3). Its overall scores were generally better than Group 3 (who had more scores above threshold competence, but also more scores below threshold competence). These results are not statistically significant, however.

Group 3 compared with other groups

As noted above, Group 3 was significantly poorer at giving advice at the right time than Group 1 (about 10% of cases were poorer). They were also less likely to incur disbursements where disbursements were actually appropriate than Groups 1 and 2.[44] Group 3's failure to carry out other work was felt to be inappropriate in a greater proportion of cases than Group 2. This may provide an indication of how the constraints of contracting work against Group 3;discouraging disbursements and the carrying out of appropriate work.

When compared with the NFP agencies, Group 3 did less well in a number of areas. These were: effectiveness of fact and information gathering;[45] legal correctness of advice;[46] appropriateness of the advice to the client's instructions;[47] comprehensiveness of advice;[48] and whether the advice was given in time/at the right time.[49]

Further fact finding was generally less likely to be appropriate[50] and efficiently executed.[51] Similarly, other work was also less likely to be

---

[44] Both these criteria were ones where peer reviewer variability means the results should be treated with caution.

[45] Chi-square p = 0.035

[46] Chi-square p = 0.015.

[47] Chi-square p = 0.050

[48] Rated negatively in 43% of Group 3 cases compared with 29% for NFPs, positive ratings were 28% compared with 37% for NFPs, p = 0.007.

[49] This did not occur in 15% of Group 3's cases compared with 5% for NFPs, p = 0.001.

[50] 23% of cases were rated negatively compared with 11% for NFPs and 30% were rated positively compared with 46% for NFPs, p = 0.001. This is a criterion where peer reviewer variability means the results should be treated with caution.

[51] With 26% being assessed negatively compared with 12% for the NFP sector. Positive assessments were at 31% compared with 38% for NFPs, p = 0.002. This is a criterion where peer reviewer variability means the results should be treated with caution.

appropriate[52] and efficiently executed.[53] Effectiveness in achieving what the client wants from the use of letter-writing and form-filling was poorer.[54]

Group 3 did less well at informing the clients of the merits of their case[55] and keeping them informed of developments.[56]

Group 3 also failed to consider, advise and/or act on referral in 64% of applicable cases compared with 35% for all NFPs.[57]

The overall score for Group 3's files was also significantly lower than for NFPs, with 33% of cases being assessed negatively compared with 25% of NFP cases and 27% of cases being rated positively, compared with 42% for the NFP agencies' cases. This suggests that upwards of 1 in 5 clients got a better overall quality of service from NFPs than from Group 3 firms.

Beyond these statistically significant results, the broad trends indicated that Group 3 performed slightly more poorly than all the other Groups. Interestingly this appeared to be partly caused by a greater diversity of service with Group 3 files assessed at above average levels of negative scores and slightly above average levels of positive scores in general. Unfortunately Group 3's negative scores significantly outnumbered the positive scores. This may be a sign of Group 3 shifting its approach to adapt to contracting in its stronger form and taking a more strategic approach to its work (filtering cases which need to be handled well). Equally, it may due to more diverse levels of performance within Group 3 (some handling the new contracting regime well and others less so).

What else drives quality (as assessed by peer reviewers)

A number of other factors were also likely to influence the outcome of peer review. Assessing these factors helped to provide an indication of how much difference in quality (as assessed by peer reviewers) was due to groups and how much to other factors (including variation between peer reviewers). It also provided an interesting indication of the extent to which quality is driven

---

[52] Chi-square p = 0.001.

[53] 25% of cases were rated negatively, compared with 15% of NFP cases. 31% of cases were assessed positively compared with 42% of NFP cases, p = 0.008

[54] 29% negatively rated compared with 22% for NFPs and 27% of cases rated positively compared with 47% for NFPs, p = 0.001.

[55] 35% of cases assessed negatively compared with 23% for NFPs and 27.6% assessed positively compared with 38% for all NFPs, p = 0.007

[56] 30% of cases assessed negatively compared with 23% for NFPs and 25% assessed positively compared with 44% for all NFPs, p = 0.003.

[57] p = 0.004. This is a criterion where peer reviewer variability means the results should be treated with caution.

by factors identifiable within the BriefCase database. As a result, other variables were investigated to see if they had any significant effect on the distributions of overall scores under peer review (on the three point scale). The following variables were considered:

> gender, marital status, ethnicity, subject category, legal aid area office (Legal Services Commission region);
>
> what level the case was handled at, by only experienced solicitors (or equivalent), only qualified solicitors (or equivalent), or only trainee solicitors (or equivalent), or by mixed levels of adviser; also the proportions of time spent by each level of fee earner were considered;
>
> whether or not there was a positive financial result; the identity of the peer reviewer; and the contractual group.

The following factors were identified as apparently affecting peer review scores:[58]

> There was a trend ($p = 0.063$) towards the different profiles for different subject categories in terms of peer review assessment and quality. Thus, debt and personal injury cases were rated more highly than welfare benefits cases. Housing cases and employment cases were the most likely to be poorly rated by peer reviewers.
>
> The contractees geographical location had a significant effect on its peer review results.[59] To focus on the four main areas in the pilot, Liverpool had 35% of cases assessed as below threshold competence; in London the figure was 21% of cases; in Nottingham 20% of cases and in Leeds 15% of cases. Cases were assessed as being above threshold competence in 38% of cases in London; 37% of cases in Leeds; 29% of cases in Liverpool and 18% of cases in Nottingham. This is evidence of a regional variation in quality. Leeds appeared to be scoring more highly in relation to peer review cases than London, Nottingham and Liverpool (in that order). This was not explained by particular peer reviewers being allocated to particular areas.[60]
>
> The peer reviewer assessment of quality was significantly different when the level of adviser handling the case changed (even though the reviewer would not have necessarily been aware of the level of any

---

[58] It is only a prima facie case because of the need to assess the extent to which such variation is caused by the variable under investigation or some other variable associated with it.

[59] (Kruskal-Wallis test, $p = .00011$.)

[60] There was a healthy spread of peer reviewers going to different areas of the country. A multinomial regression which compared only area of the country, peer reviewer and peer score still found that the geographical area itself significantly affected the overall mark under peer review when the identity of the peer reviewer was controlled for ($p = .0001$).

particular adviser working on a matter).[61] Where cases were handled only by experienced advisers, then cases were likely to be judged as being of a higher quality under peer review (fewer below threshold competence and more cases above threshold competence). Similarly, cases which were only handled by advisers of the level "trainee solicitor (or equivalent)" were more likely to be judged below threshold competence (38% compared with the average for all cases of 25%) and less likely to have cases assessed as above threshold competence. Both of these results suggest that the level of experience of the adviser had an important bearing on the quality of the work and/or the peer assessment of the file. Similarly, cases that were only handled by qualified solicitors (or equivalent) tended to get slightly poorer scores than experienced solicitors. Cases where there were mixed levels of adviser working on a case, had similar results to those handled only by experienced caseworkers.

Cases where there was at least one positive financial result were significantly more likely to be rated highly by peer reviewers. 59% of cases where there was at least one positive financial result were rated as above threshold competence compared with 28% of all cases. This was a strong effect.[62]

The identity of the peer reviewer had a significant effect on the outcome of the peer review on each file.[63] In particular PR3 had a higher "failure" rate than other peer reviewers. This reflects the variability highlighted earlier in the Chapter. The regression controls for this effect when assessing the relative impact of each factor in regression calculations.

Files all had significantly higher scores under peer review if they had more time spent on the first meeting; further work; putting the client's case; and (although very small average times were dealt with here) contractual representation; and more total time spent on a case.[64] This suggests either that it takes more time to do a better job or that peer reviewers were more willing to grant better peer review scores to files where they could see a lot of work had been done.

Another relationship that was investigated was the relationship between peer review assessment of quality and the length of time the case took. Cases which were assessed as below threshold competence tended to take the longest time (an average (mean) of 107 days). Conversely, cases which were assessed as threshold

---

[61] (Kruskal-Wallis test, p = .00011.)

[62] Mann-Whitney test, p = .00011.

[63] Kruskal-Wallis test, p = .00011.

[64] Anova tests, p= .00011 save for total time spent on contacted representation, p = 0.011.

competent tended to take 85 days. Cases which were assessed as above threshold competence took an average of 96 days. These results were significant and suggest an interesting phenomenon. Peer reviewers may be distinguishing between cases which are taking too long through inactivity and delay (which are rated as below competent) and cases which need a lot of time and effort (which are then assessed as above competent as a result of that fact) as well as the more "standard" bread and butter cases which take a shorter length of time and which are assessed as neither below or above threshold competence.

As already discussed above, there were significant differences between the distributions of overall scores between the contractual groups.[65]

A multinomial regression looked at the impact of each of these variables on the peer reviewers overall score of each file (three-point scale). A regression enables some control for variation in each of the other variables to see how much of the variation in a peer reviewer's score is independently attributable to a particular factor. So, for example, the ability to control for such variation enables the research to separate out the variation caused by peer reviewers themselves from variation caused by group and other objective factors.

All factors which have been identified above as showing significantly different distributions for peer review overall score were entered into the regression calculation, save where the variables duplicated each other (e.g. amount of time spent by experienced solicitors and cases where only fee earners at the level of experienced solicitor worked on files).[66]

Table T5.5 sets out the significant (and near significant) coefficients indicating which factors were most likely to have an independent influence on the peer review scores. As a result, it appears that the following factors were most likely to influence the outcome of peer review:

Where the case was handled by a solicitor, rather than an NFP agency, the likelihood of a case being assessed as below threshold competence increased markedly (and conversely such cases were far less likely to be assessed at above threshold competence). Differences between the three solicitors' groups were not significant when other independent factors were controlled for.[67]

---

[65] Kruskal-Wallis test, p = .00011.

[66] The variables entered into the regression were: total time spent under contract on a matter (in hours); level of adviser working on a matter; case length (in days); work category; LSC region; the existence or not of a positive financial result; the contractual group a contractee was in (i.e. group 1, 2, 3 or NFP); and the identity of the peer reviewer.

[67] Although the coefficients suggest that Group 2 actually may have been performing slightly more poorly than Group 3; when independent factors were controlled for this difference was small and not significant.

Similarly the presence of positive financial results had a significant impact on the peer review score, with the absence of such results increasing the likelihood of the file being assessed at below competence;

An increase in the amount of time spent per case had a positive impact on the assessment of quality reducing the likelihood of an assessment of below competence and increasing the likelihood of an assessment above threshold competence;

Three of the peer reviewers had a significant independent impact on the peer review scores: PR3 was less likely to assess cases as competent than the other reviewers,[68] whereas PR2 and PR6 were more likely to assess cases as above threshold competence and less likely to fail cases than other reviewers;

Shorter cases (those that took up to 99 days) were more likely to be positively assessed than cases which took longer;

Personal injury cases were more likely to be rated highly than other work categories and housing cases were less likely to get marked as threshold competent than other areas (and tended to get higher marks on the whole);

Contractees based in Nottingham, Leeds and London tended to score significantly more highly than Liverpool;

Performance of contractees under peer review

Table T5.3 looks more generally at the performance of individual contractees by looking at mean scores for each contractee. This indicates that individual contractees from each group fall into each level of performance and that, in particular, a significant number from each group fall into the lowest quartile, where average marks were considerably below 'threshold competence' across a range of files. In two contractees, all files (ten in each organisation) were rated as below threshold competence.

A simpler summary indicates how these figures break down by group.

7 out of 14 (50%) of Group 1 contractees were assessed on average as being above threshold competence.

6 out of 11 (55%) Group 2 contractees were assessed on average as being above threshold competence.

4 out of 13 (31%) of Group 3 firms were assessed as being above threshold competence.

---

[68] This result was not quite significant, p = 0.064.

9 out of 14 NFP agencies (64%) were assessed as being above threshold competence on average.

These figures emphasise the stronger performance of the NFP sector but also suggest that all groups have their poor and well-performing organisations. The following table provides the safest assessment of whether a contractee has passed or failed a peer review visit. It distinguishes between those contractees which have mean scores which are significantly above and below threshold competence, as opposed to scores which are not statistically significantly different from threshold competence.[69]

Table 5.7: Peer review assessment of contractees

Group	Above threshold competence		Threshold competence		Below threshold competence	
1	2	14%	10	71%	2	14%
2	2	18%	7	64%	2	18%
3	2	15%	8	62%	3	23%
NFPs	6	43%	5	36%	3	21%
Total	12	23%	30	58%	10	19%

The results from this table present an interesting and significant emphasis to earlier assessments of overall quality and differences between NFP and solicitor cases.

First, we can be confident that about 1 in 5 of the contractees were performing at levels below threshold competence.

The second lesson is that the proportion of failing organisations is similar in the NFP sector and the solicitor groups (and possibly worse than Group 1 and 2).

Thirdly, the number of NFP contractees performing at higher levels of quality is significantly higher than for the solicitor contractees. As a result, in terms of contractees performing at higher levels of quality, the NFPs in the pilot clearly outperformed the solicitors.

Summary of Peer review Assessment of Cases

Peer review generally assessed performance on cases as satisfactory. There were some areas of specific concern (e.g. referrals) and some indications of economic incentives inhibiting work from being carried out or disbursements being incurred which, in the view of peer reviewers was inappropriate. In

[69] A one-sample t-test was performed on each contractee's mean to indicate whether, at a 95% confidence level (i.e. $p < 0.05$), the contractees mean score was significantly different from 2 (threshold competence) on the three-point scale. As will be seen from Table T5.3 a large number of contractees that had peer review scores that averaged below 2 were given the benefit of the doubt on this test, although their level of performance gives rise to some concern.

40% of cases where no further work was carried out beyond initial advice, peer reviewers regarded the failure to carry out further work as inappropriate. When assessed on the legal correctness of advice, 22% of cases were rated as below threshold competence.

More notably, when contractees were assessed across a number of cases, 1 in 5 pilot organisations were generally performing at a standard significantly below threshold competence. For these contractees poor performance was not an isolated phenomenon restricted to one or two cases.

Importantly, and contrary to the main thrust of this research, the number of NFP organisations performing at an inadequate level was similar to the solicitors. In terms of levels of poor quality, NFPs and solicitors appeared very similar. Conversely, in terms of contractees performing at higher levels of quality, the NFPs in the pilot clearly outperformed the solicitors.

The differences between the three solicitor groups were generally not statistically significant. That said, evidence pointed towards Group 3 being the poorest performer of all.

This study also demonstrates a number of factors which influence quality especially the length of time spent on a case and the achievement of financial results for the client. These results also underline the importance of being aware of other variations in quality. Different areas of the country, and different types of work seem to have different cultures and levels of quality. Similarly, there is an important element of subjectivity in peer review assessments of quality demonstrated by the variability caused by different peer reviewers. This emphasises the need carefully to train and monitor peer reviewers and also adopt careful analysis of results.

Model Clients (Chapter 7)

This chapter provides a detailed analysis of model client and peer reviewer comments on the model client visits. Forty-five model client visits were scheduled, covering nearly a third of contractees. The forty-five model client visits were evenly distributed between NFPs and solicitors groups, geographical areas and the four work categories looked at.[70] Chapter 2 discusses the methodology of the model client programme in detail.

Whilst model clients (or "mystery shoppers") have been widely used in consumer research,[71] and also some medical studies,[72] there has been no

---

[70] Group 1 had 13 visits scheduled, Group 2 10 visits, Group 3 12 visits and NFPs 12 visits. 14 visits were scheduled in Leeds, 11 in Liverpool, 13 in London and 7 in Nottingham. 10 visits were scheduled in debt, 15 in employment, 12 in housing and 8 in personal injury.

[71] For a discussion of the benefits of 'mystery shopping' see Consumers Association (2000) *The Community Legal Service: Access for All? Policy Report*, (Consumers Association, London, 2000) p. 18.

controlled use of model clients in socio-legal research on this scale.[73] Given the significant number of such visits, the programme reveals important qualitative insights into the general levels of quality of service provided under contracting and specifically uncovers examples of poor service.

There are two codas to this observation. The model client visits assessed only the early aspects of service: access and the quality of advice and immediate follow-up at the initial interview. These are, of course, crucial aspects of the service as will become apparent in this chapter. Advisers' later work in handling cases to completion was assessed by other methods and is discussed elsewhere in this report. Secondly, these visits represented only one visit, (usually) to one adviser within each contractee. As such, model client visits were not a meaningful way of assessing the range of quality within individual contractees. They do, however, provide a useful assessment of initial work under contracting.

The chapter begins by discussing access problems experienced by the model clients. It then looks individually at each model client scenario (the debt model, the employment model and the housing model) to provide a detailed understanding of the level and quality of advice given under contract. The chapter ends by conducting some statistical analysis on quantitative aspects of the results. Model client visits have been noted by a unique number and the work category within which the visit fell. Peer reviewer comments have been labelled by a unique peer reviewer number (PR 1 to 6).

Access problems under contract

The model clients reported a number of difficulties in accessing the service. These fell into four main types:

Difficulty making initial contact with the contractees;

Difficulty getting contractees to make appointments;

Difficulty getting contractees to give advice.

Difficulties in the manner in which advice was given.

Five visits did not take place at all because of access problems. All five instances of failed visits were with solicitors' firms amounting to 5 out of 33 (15%) of scheduled visits with solicitors. This was usually due to contractees stating that they did not do that work or to persistent difficulties in making

---

[72] See, for example, Rethans, J., Drop, R., Stumans, F., Van der Vleuten, C. (1991*) Assessment of the Performance of General Practitioners by the Use of Standardised (simulated) Patients*, (1991) 41 British Journal of General Practice 97-99.

[73] Wasoff, F., Dokash, R., Marcus, D. (1990) *The Impact of the Family Law (Scotland) Act 1985 on Solicitors' Divorce Practice* (Control Research Unit, Scottish Office) being closest to a model client study.

appointments. In one of these scheduled visits the contractee did little of the relevant work category. In the other four, however, each contractee had closed at least forty-five matters in the relevant work category. It may be that these 'failed visits' were a response to short-term pressure of work. Even where this occurred because of inability to do the type of case, clients were not usually referred on to other suppliers.

All scheduled visits to NFPs took place but in 5 out of 12 scheduled visits (42%) significant access barriers were placed in the way of the model clients. The following model client report illustrates some of the initial difficulties in making contact with some contractees and then, once contact was made, getting to see someone who might be able to advise:

> "It was difficult to arrange to see someone here – the answer phone was on a lot of the time or the line was engaged. When I finally got through I found that it was a drop in centre, open from 9.30 to 3.30. I arrived at 12 noon and was asked to take a ticket and as I would be unlikely to see anyone until 2.30 p.m. I should go away and return later. I didn't get seen until 2.55 p.m." (Employment38, NFP (CABx))

These problems were particularly apparent in the CABx contractees. There were a number of examples where model clients were asked to come back at different times, and required to wait for significant periods of time. Sometimes, the model client had to be very persistent indeed if they were to get access to the agency:

> "Problems first arose when I arrived at the CAB at 13.10 and found it closed although having telephoned the previous day I had the impression from the answer phone that it was a drop in centre up to 2.00 p.m., with no appointment necessary. As I turned away I noticed that lights were still on and people were inside. I rang the bell and was asked if I had an appointment. I explained that I had rung up and heard the answer phone advising me that I didn't need one and that I needed to see someone urgently today and was unable to return tomorrow. I waited ages before being let in. The person waiting with me who did have an appointment said to me "They're always closing early". Neither the person who eventually opened the door nor the receptionist was friendly. They both seemed put out by my persistence and the receptionist barked at me "you don't have an appointment do you." It seemed not everyone waiting there had an appointment time. I waited two hours to be seen by an adviser who seemed surprised that I had waited so long." (Employment 42, NFP (CABx)).

Such problems were not confined to drop-in approaches to advice delivery. In another example the model client persisted with three telephone calls before advice was given which they felt was reasonably comprehensive. The first telephone call was met with, "brief advice in response to my brief portrayal of problems. Told they didn't make appointments to see anyone in person unless absolutely necessary after phone advice." A week later, a second call was made where there was an answer phone message apologising that no one was available. On the third telephone call, the model client, "spoke to same adviser again. She remembered me and had some notes." Advice was then given on the telephone and a follow-up letter sent. (Employment 39, NFP).

Once through the door, however, problems with the service did not end. On at least two occasions the model client saw a succession of people in circumstances which they found confusing and which also risked compromising their confidentiality. This example illustrates the point:

> "On arrival – one person checked my name off and I sat down to wait. A few minutes later – a second person took some details from me at a desk quite close to waiting area. I had to sign a form to say that I understood the type of service they were providing, then I had to go back to waiting area and wait again to see someone else. This was not explained to me until I asked what was going on. Very confusing that I had to see two different people as well as having been asked to provide written details of my problem myself and bring these to the appointment." (Employment 32, Group 1 of Solicitors Pilot)

On seeing a third person (a solicitor) the client was finally advised. One of the peer reviewers commenting on this model client report noted, "The system used by this organisation is clearly deficient – it does not provide confidentiality and the explanation of the problem to three people is inappropriate." (PR 4).

These access problems were a significant feature of a large number of model client visits and were largely confined to NFP agencies. To an extent such problems appeared to be caused by overloaded agencies unable to cope with the demands on their services. Where a client was passed from a receptionist to a generalist (possibly volunteer) adviser and then, either directly or via the generalist adviser, to advice from a more specialist worker, this illustrated how the service was structured in some agencies. Such a structure aims to protect the time of the specialist workers to deal with cases requiring their expertise. Such structures, as well as giving rise to access and service problems, may be one cause of the inadequate advice indicated below. Equally a small question mark should be registered over the need for shielding contracted workers in this way, given that many NFP agencies had, during the life of the contract, reported difficulties in meeting their 1,100 contract hours (or higher targets in some agencies).

The above examples came from the employment visits, significant problems were also found in the other work categories. Solicitors firms were not found to demonstrate the types of problems exhibited by NFPs, but did decline to make appointments for a number of clients. 2 out of 10 debt model clients were turned away by solicitor contractees without advice. One contractee did no debt work, but the client had to be quite demanding before they got a referral.[74] The other debt visit was abandoned after the model client telephoned persistently without being able to speak to anyone. 2 out of 12 housing visits experienced similar problems and so no advice was given. In addition to the above problems, 1 of the scheduled employment visits did not take place because of access problems.

---

[74] "[I was told,] No-one deals with debt. This time I had to push quite hard for any further help and was given the telephone number of …..another law firm." (Model Client Visit 4, Group 2 firm).

Model clients and the quality of advice

In addition to the model client's own assessment of the service on questions of access, and further questions discussed below, the model clients completed a written report on the visit. They were asked to set down exactly what the adviser had told them about their problem and, in particular, what the adviser said about:

The exact problem;

What their rights in law were;

What the model client could do about it;

What the adviser can/would do;

What, if anything, the adviser was proposing to do next;

Whether the adviser would be confirming their advice in writing; and,

Whether the adviser considered that they wanted more information from anybody else.

These reports, along with any follow-up correspondence sent by the contractee, were then assessed by the peer reviewers who marked the model client visits on the same five point scale as used during the peer review exercise and provided written reports on the quality of work. These written reports are analysed in this section for each model (debt, employment, housing and personal injury). Quantitative analysis of the marks is dealt with at the end of the chapter.

Debt model

In the debt model, the client had [recently] been made redundant after a few months employment. The client had a consumer credit agreement on a TV and video that committed him/her to regular monthly payments of £28 pcm over a year. The client was worried about their ability to meet these payments. The model client was instructed to ask the contractees what their legal position was and, in particular, whether they should try to negotiate with the creditor, terminate the agreement or simply let the debts build up. The model client did not take the written credit agreement to the adviser for the interview.

In broad terms, peer reviewers indicated that contractees might be expected to advise that if the client breaches the agreement by failing to pay the contractual instalment, the likelihood was that the creditor would sue and the court would order, having looked at the model client's financial position, the appropriate payment to be made to the creditor. Similarly, they would be expected to advise on the possibility that the client would have to return the goods and pay half the price of the goods (if it was a hire purchase agreement or conditional sale agreement rather than a consumer credit agreement). Of particular importance was the need to make clear to the client that the adviser had to see a copy of the credit agreement to be able to advise on the legal

position with accuracy. The adviser would also be expected to give at least some consideration to the welfare benefits that the model client was receiving.

The adviser might also advise on the effect of a breach of contract on the client's credit rating and the effect of a county court judgment; the possibility that the agreement was invalid (e.g. because of failure properly to execute the agreement under the Consumer Credit Act 1974); and, the possibility of payment protection insurance in case of redundancy (although that would be unlikely given that the client had only been in employment a few months).

There were other aspects of service which could be considered, probably beyond the first interview,[75] but the main thing the adviser needed to convey was that there are important consequences to the client if they simply let the debt build up.

Debt model client visits

Ten suppliers were visited by a debt model client. Five advised correctly and three advised in a way that peer reviewers regarded as either incorrect or significantly incomplete. Two turned away the model client without advice. The following comments relate to the three visits where advice was given but was regarded as deficient by the peer reviewers.

In the first problem visit (Visit 6 (Debt) NFP with debt and welfare benefits contract), peer reviewers' concerns centred mainly on the incompleteness of the advice. The adviser did not ask to see the agreement, or explain that sight of it was necessary to advise accurately. The advice concentrated on a practical solution (negotiation of terms) but did not discuss the legal position, save to say that the court might rule if the model client was sued by the creditor. There was no attempt to ascertain the model client's income and expenditure, nor any welfare benefits advice. No follow up letter was sent to the client.

Interestingly the model client was quite happy with the service, but the peer reviewers were less impressed:

> "No advice on possible termination rights (the model client particularly wanted to know whether she can terminate the agreement) and no request for sight of the agreement. Incomplete advice." (PR2)

> "Did not see or ask for an agreement, therefore incomplete or inaccurate." (PR1)

[75] A financial statement should have been prepared detailing income and reasonable expenditure. The amount of the excess would indicate the levels of payments that could be sustained. Advice should also have been given to the client on priority and non-priority debts. It is questionable whether the adviser should have gone into great detail on the different types of advice appropriate to different types of consumer credit agreements. These could have been outlined, and more advice given when the client brought in the agreement and the financial statement was prepared.

In the second example (model client visit 7), the incompleteness of the advice appeared to be more problematic. The adviser did not explain the model client's legal rights, there was no request to see the agreement, or any investigation of the model client's finances (beyond the legal aid eligibility test). The model client was led to believe both that he had no rights and that neither he nor the adviser could take action until a court claim against him was commenced by the creditor. The adviser suggested the model client write to the creditor but offered no advice on content or wording. No welfare benefits advice was given. Similarly, no follow-up letter was written in spite of this being a contract requirement for this contractee (Solicitors' pilot, Group 3 contractee).

The model client was not impressed by the level of service received from this contractee and showed an inkling of the consequences of the lack of proactivity:

> "I may well enter into a period of debt. I have very little recourse to the law as I have broken an agreement. The adviser told me she could do nothing until I have actually run into debt and court proceedings were planned. Wait and see what develops."

Similarly, the peer reviewers had concerns about the quality of work carried out as the following comments indicate.

> "Very little advice other than send a letter and offer a pound. Basically come back when you are really in trouble then we'll see what we can do. No sample letter or positive assistance."(PR5)

> "Very little advice. No request for further information. Advice given not very practical and unlikely to materially improve client's situation." (PR1)

> "The advice appears to have been wrong as well as incomplete. The adviser effectively tells the client that s/he has no rights 'as I have broken an agreement' and that there is nothing the adviser could do 'until court proceedings are planned'. Incomplete, inaccurate and misleading advice."(PR2)

> "Incorrect advice. The adviser cannot not do anything until the court proceedings are planned. Inaccurate and incomplete."(PR3)

The third example is perhaps the worst. In model client visit 8, the model client was advised that she had breached the agreement and so had no legal rights. There was no request to see the agreement, nor an attempt to ascertain the model client's finances. No welfare benefits advice was given (in spite of this being the work category in which the agency had a contract). A letter to the creditor was suggested, but the content was not specified. No action was proposed by the adviser nor was a follow-up letter sent to the client. (Model Client Visit 8 (Debt) NFP agency with welfare benefits contract**)**.[76]

The model client had similar views to visit 7. The reviewers note:

---

[76] This contractee did not have a debt contract but had carried out some debt work under its contract.

> "It is simply wrong that the client has 'little rights as it is my fault for not being able to pay money'. Incomplete and inaccurate advice." (PR2)
>
> "Very little advice, no request for further information e.g. seeing agreement, overall debt situation. Little detail on how to approach creditor."(PR1)
>
> "Model client had no rights and could be made to pay money she didn't have. Very poor advice and wrong. (PR5)

Two of the visits where problems were identified were with contractees that did not do significant amounts of debt work. Whilst this may explain their performance, it is interesting to note that, rather than acknowledge the absence of knowledge in this field and refer the client the adviser proceeds to give poor advice to the client.

Where there were problems, the debt model tended to reveal incomplete advice and advice which was inaccurate. Even visits which the peer reviewers regarded as "passing" tended not to get higher marks on the scale. As the model client visit 8 showed, this incompleteness could significantly undermine the client's position in dealing with the creditor and had the potential to make matters worse for the client.

Housing Model

In this scenario, the model client was a tenant who had their gas boiler disconnected by British Gas because it was unsafe. British Gas had been called out by one of the tenants in the building who smelt gas. As a result, the tenants had no hot water and no heating.

The model client had approached the landlord to carry out the repairs. The landlord appeared to be insisting that one of his regular workmen who was not CORGI registered should carry out the repairs himself, even though a gas plumber must be CORGI registered to do this work. The model client realised that this might result in British Gas refusing to reconnect the supply, as the repair had not been properly carried out.

The model client/tenant was instructed to ask whether they were legally entitled to get British Gas to carry out the work and then bill the landlord, or to offset the cost of the repair against the rent. They were also instructed to indicate that they wanted to deal with the landlord direct rather than inflame the situation with solicitors' or advice agency letters.

In this situation it was the view of the peer reviewers that the landlord has a duty to provide services to the tenant which are necessary for occupation of the premises as a home, e.g. gas.[77] The landlord is also required to ensure that a CORGI registered contractor carries out an annual check on the boiler. If these checks are not carried out the tenant has recourse to the Health and Safety Executive. The HSE can enforce safety requirements, although this can

---

[77] Section 11 of the Landlord and Tenant Act 1985 imposes a duty on the landlord to keep in repair gas pipes and fixed heaters etc.

also come under the Environmental Health Department depending on the nature and extent of the disrepair.

The legally correct and complete answer would explain that the landlord has a duty to provide gas, a duty to keep gas pipes etc. repaired, and an annual duty to carry out safety checks on the gas boiler using a CORGI registered contractor. If the landlord appeared unwilling to carry out the repair or was proposing to have it done by someone not suitably qualified, then the tenant should have considered giving formal notice to the landlord that he would have the repair carried out by a suitable contractor, and obtain three estimates prior to doing so. The adviser should warn of the dangers of withholding rent and consider the possibility of involving the HSE/EHD.[78] The tenant should also be advised of the possibility of a damages claim against the landlord. Given the need for a practical solution, the model client/tenant should be advised of the best negotiating strategy and the legally correct fallback position should the negotiations fail.[79]

Housing model client visits

The housing model provides a good example of the difference between advice which identifies a breach of a legal duty and advice which goes on to provide the client with practical assistance to remedy the problem that they face.

In one visit, apart from stating that the landlord was obliged to provide safe premises and heating, and that the lack of these was 'serious', the adviser provided little legal information to the model client. There was no explanation of the legal framework, nor of whether the model client could offset rent against the costs of repairs. The follow-up letter simply reiterated the adviser's willingness to write to the landlord if the tenant so instructed (Model Client Visit 20, Group 1 firm).

The following are a sample of the reviewers' comments on this visit:

> "Basic….accurate….incomplete. Follow up letter – brief – does not give any advice simply restates the problem."(PR1)
>
> "My assessment is that this solicitor/adviser does not know much about housing law. There is no precise advise about implied repairing covenants or the possibility of the EHD being involved. The letter … setting out the advice given is equally bald and uninformative…"(PR4)
>
> "The advice was incomplete in that only general advice was given which I would call basic but accurate."(PR5)

---

[78] The peer reviewers' opinions were split on the merits of withholding rent. Some thought this could be dangerous as it provides the landlord with cause of action against the tenant, others that there was a right to set off sums correctly spent on repairs against any rent due, provided due notice is given to the landlord and three estimates obtained for the work.

[79] There was the possibility, depending on the outcome of negotiations with the landlord, of the tenant seeking specific performance and possibly damages.

The importance of explaining to the client in practical terms, what legal and other remedies are available is worth stressing. Clients who wish to go and negotiate with landlords need some idea of what they can compel the landlord to do (and what the implications of that might be). A similar problem is evident in the next model client visit. Here, the model client was given quite full advice on their rights[80] but little practical advice on how to pursue those rights. As the model client reported, the adviser described the situation as 'not good' and that a discussion between the tenant and the landlord should be their next course of action. No further action was proposed and no follow-up letter was sent (Model Client Visit 22 to a Group1 firm).

The model client had concerns about this adviser because they did not advise them on any action and did not appear interested in the problem. Similarly, the model client also did not feel they were given enough time to make relevant points to the adviser. The following comments capture the peer reviewers' opinion:

> "Accurate as far as it goes but very incomplete. No advice on action to be taken."(PR1)

> "No advice given as to basic rights, foundations for them or remedies to be taken in the event of breaches."(PR5)

Similar problems were found in other visits.[81] Peer reviewers criticised advisers for giving woolly advice, failing to offer practical strategies to require the landlord to take action, and failure to address the specific questions raised by the model client. It is notable that several of these failures took place in CABx, where the advisers (especially if generalist advisers) may have little experience of taking action for clients. Showing concern and having a basic understanding of the client's rights may not be enough. The following model client reports one adviser's words of wisdom:

> "[S]ometimes landlords can be difficult to deal with. The problem needs to be resolved as it is very cold at the moment."

However, the adviser did not advise on any concrete course of action.

5 out of 12 of the housing model client visits contained wrong or incomplete advice and two failed to advise for similar reasons to those seen in the debt cases. This left only five providing advice that the peer reviewers felt was correct and reasonably complete. The next section, which looks at the employment model, suggests even stronger levels of concern.

---

[80] They were advised that they had a right to heating and hot water and that an annual CORGI appliance appraisal was an entitlement.

[81] For example, Model Client Visit 21 to an NFP (CABx) with housing and welfare benefits contracts, Model Client Visit 27 to an NFP agency (CABx) with a housing contract.

Employment model

In this scenario the model client was employed part-time without a written contract as a waitress. She had money deducted from her wages without her agreement because money had gone missing from the till. The employer had decided to penalise all the employees equally for this "theft", for which the employer could identify no perpetrator. The Wages Act provides that no deduction may be made from an employee's wages without their prior written consent.[82]

A related concern was what would happen if, the client raised the question of an illegal deduction with the employer, after which the client was then dismissed. A crucial point was that, although the qualifying period of employment for taking an unfair dismissal claim to an industrial tribunal was (at the time of the visit) two years, if the employee is asserting a statutory right,[83] a claim can be brought without any qualifying period.

There are other subsidiary issues that may arise. There is an entitlement to an itemised pay slip, and a remedy if none is given. An employee is also entitled to a statement of terms and conditions of employment.

A correct and complete answer would deal with the general principle that no deductions can be made from an employee's wages without their prior written consent and consider the potential exception under the ERA 1996. It should then go on to deal with what the client could do if dismissal follows; the possibility of making an application to an Employment Tribunal if dismissal arose from asserting a statutory right. Equally, the adviser may need to raise the possibility that the employer would argue dismissal on other grounds (e.g. incompetence).

It should be remembered that the client had not been dismissed and was seeking advice on how to tackle the issue with her employer, ideally jointly with the other workers.

The second set of model client visits were carried after the introduction of the minimum wage. The client was being paid below the minimum wage. This is a criminal offence on the employer's part and an application could be made to an Employment Tribunal seeking an order that the employer pay the minimum wage and arrears. On the second set of model client visits the advice should have covered this also.

A number of contractees got the legal advice wrong and advised the client that there was little that they could do in the situation that they found themselves.

---

[82] Although there is an exception that applies to retail workers under Sections 13 to 17 of the Employment Rights Act 1996 it is debatable whether it applies in this scenario. If it did apply the employer could deduct up to 10% of the client's gross wages for theft and/or missing stock.

[83] It is a statutory right not to have money deducted from your wages. It is debatable whether the exception in S.17 applies to the facts and the client could, if they assert their rights in good faith, gain the benefit of this statutory protection from dismissal (ERA 1996 S.108 (3)g and section 104).

Thus on one visit, the client was advised that to dock pay was illegal, but that if sacked there was no recourse to the Industrial Tribunal, as the client needed a qualifying two year employment to bring an unfair dismissal claim. This advice was confirmed in writing. It concludes that the model client has no 'real legal redress against your boss…' (Model Client Visit 31, to a Group 2 firm). This advice was in notable contrast to the "woolly" approach to the debt scenario that called for criticism. Peer reviewers expressed concern that the advice in this case was confident but wrong.

In another visit, the adviser advised that if the client was sacked then no legal protection was available as they had not been employed for two years. The adviser also advised that dismissal for complaining about deductions might provide the basis for a tribunal claim, but was unsure and made no attempt to verify or clarify the advice subsequently. When the client asked for a written explanation, she was given copies of the statutes the adviser had consulted which included the point about asserting statutory rights, although the adviser had not addressed it clearly with the client. The letter confirming the advice to the client was also ambiguous: it did not appear to exclude the possibility of an Employment Tribunal claim, nor does it state that one exists (Model Client Visit 32, Group 1).

The peer reviewer comments complete the picture of incomplete, confusing and contradictory advice being given to the model client:

> "Incomplete, inaccurate and misleading, when it says "illegal to deduct pay, but no remedy because under 2 years' service. Follow up letter provides model client with copy documents which specifically say a claim for unlawful deductions can be brought where the worker has been employed for less than two years in certain circumstances. The third follow up letter appears to add to the advice given i.e. implies there is a remedy and mentions three months time limit from first deduction but contradicts the initial advice."(PR2)

> "These are either photocopied sections from Halsburys statutes or similar about deductions from wages. This is not appropriate for a lay person as incomplete: the adviser failed to look up the point about protection from statutory rights."(PR3)

> "Adviser concerned that the deduction of pay was illegal but clearly not in a way that was clear to the client. Did not advise on any way of getting the money back or the right to do so. The photocopy of the legislation provided is in my view daunting and unhelpful."(PR6)

On another occasion, the adviser, whilst sympathetic to the model client's situation, advised only that such deductions were probably illegal. Having recognised the limitations of this advice, the adviser suggested the client go elsewhere. The advice was given by telephone. This advice is of particular concern as the contractee in question held (and only held) an employment contract. (Model Client Visit 35, NFP agency). This was also an interesting example of the difference between expert and lay views on quality. The model client said:

> "The adviser showed an impressive level of concern for my job security, understanding that I could not afford to lose my job. Made a point of telling me that it can be quite common for part time women workers to encounter

> unfair bosses because they know how much they need the job and think that they will be able to get away with it (i.e. treating them unfairly). Overall he was very helpful, reassuring and personal and tried to think of as many other organisations I could turn to as he could. The people he suggested were a CAB, ACAS, a local law firm, a trade union."

Peer reviewers were more concerned. The following comment is perhaps most apt.

> "Although very clearly empathetic, this adviser does not really know enough about the law to be using legal aid money. A good example of touchy feely advice."(PR4)

There were other examples where the peer reviewers felt the model clients were not handled well. On one, the client was told it was illegal for the employer to deduct wages without the client's written consent, but no explanation was offered of the legal basis for this, or what remedies might be available, except for a general application to an employment tribunal. The model client was given an extract from the CAB database about deductions from wages, which states that a claim can be made in these circumstances to an Industrial Tribunal without any qualifying employment period, but this was not explained orally. The model client was able to understand the paperwork, but a client with lower literacy levels might have found this hard. No follow-up letter was sent. No mention was made of failure to pay the minimum wage, even though the minimum wage was in force at this time (NFP agency (CABx) with debt, employment and welfare benefits contracts). As well as commenting on the failure to advise about the minimum wage, adverse comment was made on the method of delivery as this quotation illustrates:

> "The advice given was basic but appalling in the delivery." (PR5)

The final example illustrates how a client wanting to solve the problem themself was handled by a solicitors firm. The client was told (by telephone) that the deductions were unlawful, and could be reclaimed via the Industrial Tribunal, although such action might result in the sack, for which no unfair dismissal claim could be brought as the client had not been employed for one year. The adviser wanted to write to the employer on client's behalf, which was declined. There was no mention of the application of the minimum wage (which was in force by then). (Model Client Visit 44 to a Group 2 firm). Having offered to write to the employer and been turned down by the client, the adviser suggested the client consult a CAB if further advice was required. No follow-up letter was sent.

The employment model scenario provided the clearest examples of inaccurate advice. In six out of fifteen visits was the client advised correctly and some were favourably assessed by the peer reviewers.[84] However, in eight visits

[84] This comment was made in relation to Model Client Visit 37 (to a Group 1 firm): "The advice given by the solicitor seems to be bang on. He correctly analyses the problem, gives the client the reassurance about her employment protection and then explains the practical problems in risking dismissal for a relatively small sum of money."

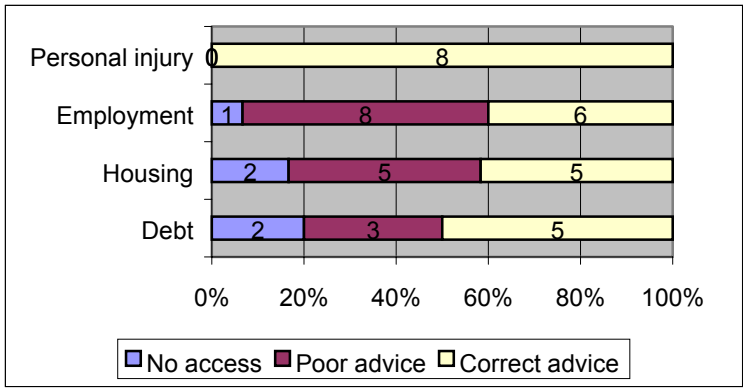
the client was advised incorrectly. In one visit there was a failure to advise because of access problems, although as the initial section of this chapter indicated the model clients on other employment visits sometimes had to be quite tenacious.

Personal injury model

Eight model client visits were also conducted in personal injury. In the scenario, the model client was involved in a car accident where their car was hit from behind whilst waiting at a pedestrian crossing while a pedestrian crossed the road. This resulted in the client receiving whiplash injuries. There was no likely sustainable dispute on liability. The personal injury scenario is different from the others in that the facts of the matter appear well established and the legal issues raised are not likely to involve serious contention or risk to the client. All eight of the model client visits were marked at threshold competence or above. There were some problems in relation to explaining things like statutory charge and other avenues of legal funding. However, broadly the comments were favourable on the advice given.

The outcome of the qualitative analysis of model client visits is summarized in the following diagram. It serves as an important reminder that when poor (inadequate, incomplete and inaccurate advice) is taken alongside inability to access certain suppliers then in employment, housing and debt levels of service were very poor.

Figure 2: Summary of model client visits (qualitative analysis)
Debt 10 visits, housing 12 visits, employment 15 visits, personal injury 8 visits.



Quantitative aspects of the model client programme and a comparison of NFPs and solicitors

As well as the qualitative report of model clients and the peer reviewers' comments on them, model clients were asked, immediately after their interviews, to complete a questionnaire evaluating the interview (Appendix C). This evaluation focused on aspects of service assessable by peer review. Comparisons between NFPs and solicitors are outlined where there are important differences. The small number of reviews means that differences are

less likely to be significant. Given these smaller numbers involved, comparison between solicitor groups would not be meaningful.

Reception

75% of organisations visited received a "yes" to the question, did the receptionist know that [the model client] was coming? Solicitors did significantly better on this (92% compared to the not for profit sector 22%) largely because NFP agencies were unwilling to make appointments for model clients who were then forced to rely on telephone and drop-in advice.[85] The model client was also asked whether, when they arrived at the contractee, the receptionist or the person receiving them made them feel welcome. This was answerable on a scale of 1 to 5 (marks of 1 and 2 indicating a poor rating and 4 and 5 a good rating) 41% of model clients rated the contractee as good, 32% as satisfactory (score of 3) and 27% as poor. Scores for solicitors and the not for profit sector were similar (solicitors doing marginally better but not significantly so).[86]

46% of model clients were kept waiting beyond their allotted appointment time. All of these waits took place in private practice settings rather than not for profit sector. In 69% of cases where there was a delay, the reason for that delay was not explained to the model client. All of these failures occurred in the private practice sector because of the virtual absence of fixed appointments in the not for profit sector.

Comprehension of the Problem

Model clients were asked if they felt that the adviser understood their problem. All contractees received at least a satisfactory rating on this criterion. 18% of cases were rated as satisfactory. 32% were rated as good or very good and 50% of contractees were given the highest rating. The profiles of solicitors and the not for profit sector were very similar.

Model clients were also asked whether they got the impression that the adviser was interested in their problem (as opposed to regarding it as trivial and/or insignificant). 11% of contractees were regarded as poor in this respect. 66% were regarded as good. There were only very marginal differences between the solicitors and not for profit sector.[87]

---

[85] Chi-square p <= .0001.

[86] Solicitors faired poorly in 27% of cases compared with 30% for not for profit sector. They scored above 3 in 42% of cases compared with 30% of cases in the not for profit sector.

[87] Slightly more solicitors were rated as poor (12% compared to 8%). 67% of the not for profit sector were rated as good compared with 64% of solicitors. These differences are not significant.

Given enough time to explain

Model Clients were asked if they were allowed enough time to make all the relevant points about their case to the adviser. Model clients felt that in 84% of matters, they were given enough time and in 16% of matters they were not. The profiles for solicitors and the not for profit agencies were extremely similar.

The model client was also asked if the adviser seemed to deal efficiently with the information that the model client gave them. In particular, whether they took notes of what the model client said, asked relevant questions and established as complete a picture as possible of the clients problem. 8% of the contractees were rated poorly on this criterion, 65% were rated as good or very good. There was a notable difference between the solicitors and not for profit sector. Both had very similar levels of poor performance (8% and 8% respectively). Conversely, solicitors had a far greater proportion of high scores (76% of their cases were rated as good compared with 50% of the not for profit sector). These differences were not, however, statistically significant.

The model client was then asked whether the adviser went on to advise them of the options that were open to them to deal with their problem. According to the model clients, all of the not for profit contractees did this. 20% of the solicitor contractees did not. This difference is not quite significant.[88]

The model client was then asked whether the adviser went on to ask questions about anything else once they had advised the model client on their problem (for example asking about and advising on whether they are claiming all available benefits). Advisers did this in 19% of cases. Interestingly, solicitors appeared more likely to do this (24% of cases) than the not for profit sector (8% of cases). These differences were not, however, statistically significant. On one occasion, the adviser had done this before they had finished advising the model clients on their presenting problem. This adviser was a solicitor contractee.

Few differences between solicitors and NFP agencies appear on these in small numbers. Solicitors seemed to deal more effectively with information, but did not advise quite as well on options. Solicitors were more likely to ask questions about other problems.

Peer review assessment of model client advice

A well as providing detailed written comments on the model client visits, peer reviewers were asked to mark the overall quality of the initial advice on a scale of 1 to 5 (as used in the main peer review exercise). This led to each model client visit being assessed by five or six peer reviewers. These marks were averaged to provide a reliable composite indication of quality of model clients' work as assessed by peer reviewers. 38% of contractees were assessed

[88] Chi-square p = .096.

at 2.5 or less. This suggests, consistent with the qualitative analysis, that a higher level of poor service was exposed by model client visits than other research methods. The average (mean) score was 2.86. Solicitors scored slightly higher (2.95) than the not for profit sector (2.3). This difference, however, was not significant.[89]

Summary

The model client data presents a useful opportunity to gain a qualitative insight into service delivery at and before the initial interview.

In the employment, housing and debt models there were significant numbers of visits where the client was either inadequately advised or incorrectly advised. In the case of the debt scenario, this appeared likely to lead to the client failing to get on top of debt problems sufficiently early to prevent problems later on. In relation to the employment scenario and the housing scenario, the advice to the client was either patchy or incomplete in a way that would probably mean that the client would tolerate inadequate housing and fail to enforce basic employment rights.

In terms of a direct comparison between the not for profit sector and private practice, the number of model client visits and the similarity in performance between the not for profit sector and solicitors meant that there are few, if any, significant statistical differences in the results. The number of visits is too small to compare between the solicitors groups.

That said, the not for profit sector did appear to be doing more poorly overall in handling these initial visits. One likely explanation is that model clients seeing not for profit agencies for the first time tended to see generalist (or volunteer) advisers rather than specialist, contracted workers. This may be one of the factors leading to the inadequate and inaccurate advice given to these clients. It should be emphasised that, even where handled by trained volunteers (not funded by the Commission), these initial contacts with advice workers can be crucial to protecting the client's future position.

Another possible explanation is that not for profit agencies did not generally send follow-up letters backing up the advice they had given.[90] They therefore did not give themselves a second chance to get it right.

A very strong piece of evidence was the frequency and level of access problems experienced by model clients seeking to get advice. The five visits (11%) which did not happen demonstrate the difficulties of real clients in

---

[89] Chi-square p = 0.249.

[90] Advice given under the Level 1 aspect of contracts did not require such advice to be given in writing. A number of comments were made by peer reviewers which suggested that the quality of written advice had increased the marks given to the model client visits. This may be one reason why the solicitors (who tended write such letters) scored more highly than the not for profit sector (who did not tend to write such letters), this is discussed further below.

gaining access to services. These failed visits all occurred in solicitors' firms. 5 (42%) of NFP visits experienced significant access problems but (through model client persistence) they usually managed to get some advice. Problems included not for profit agencies not keeping to their advertised opening hours, operating via phone lines and answering machines which did not provide any facility for leaving messages and simply requiring clients to wait for very long periods of time. There are also some concerns about referrals not being attempted when they ought to be; clients being referred in a very vague manner by exhortation to try a CAB or a law firm and clients being referred by contractees who only specialise in the work category that the client falls into.

Whilst this chapter has dwelt on the negative aspects of work conducted (justifiably so given the level and nature of problems identified), there were examples of service which both model clients and peer reviewers felt was of a very high standard. The majority of cases were graded at or above threshold competence and were praised for the accuracy, completeness and clarity of their advice. Model clients did, however, reveal a significant proportion of initial work which fell below threshold competence and what, in qualitative (or descriptive) terms such failure means. In employment, housing and debt cases the balance between poor and acceptable work was fairly even. This seemed to be true across both sectors although in general there seemed to be more failures in the NFP sector. Personal injury work was a saving grace in which all cases passed, but this should be true of all work categories.

Targeting Peer Review (Chapter 10)

How quality methods interrelate and the targeting of peer review

Chapters 5 to 9 discuss the results from individual methodologies looking at the quality of work conducted under contracting. This chapter brings the different findings together, to provide some comparison of the results from each method. In particular, the relationship between peer review and alternative quality indicators is explored. About 20% of peer reviewed contractees performed at significantly below threshold competence. This is the first time that such a detailed assessment of the quality of legal work has been carried out and it therefore provides something of a benchmark against which all future assessments may be judged. It is not objectively clear how firms and agencies were performing in the past. It is to be noted that these results occurred in spite of providers having passed through the Commission's pre-existing quality assurance mechanisms and this suggests that the further development of quality systems and indicators used by the research has been useful.

As an additional quality assessment mechanism, peer review provides a depth and subtlety of approach which does not have the limitations of other quality assurance measures. It is capable of assessing strategy and making nuanced judgements about the content of advice: and as such it is likely to be the

closest test for 'real quality' that is available. As a result, one of the main recommendations of this report is for the Commission to develop a programme of peer review as part of its quality assurance system. However, peer review needs to be carefully managed if it is to produce sufficient objectivity. It is also an expensive tool. Thus, the main aim of this chapter is to isolate factors that may help the Commission to identify poorer performing firms, or those most likely to be performing poorly, being the firms most likely to merit assessment by peer review.

Comparison of quality measures

This section compares the quality assessments of contractees using a range of techniques to examine the extent to which they might serve as alternatives to peer review or triggers for peer review.

Peer review and client satisfaction

Client satisfaction is intended to look more at the service element of the work of lawyers' and advisers', rather than the technical quality of the legal work itself. Client surveys are unlikely, therefore, to serve as a proxy or trigger for peer review assessment. This is clear when levels of client satisfaction in each contractee are compared with their peer reviewer assessments. The following table compares the client satisfaction ratings with the peer review assessments of those contractees.[91] This was possible for thirty-three contractees.

Table 10.8: Peer review and client satisfaction compared

	Client Assessment (overall)			Total (Peer Review)
Peer Review Assessment	Below average	Average	Above average	
Below threshold competence	1	4	1	6
Competent		19		19
Above competence		7	1	8
Totals (Client Satisfaction)	1	30	2	33

The results from this table show that the contractees that were assessed as above average by the client satisfaction survey were also performing at threshold competence or above according to peer review assessments. It should be remembered that the average satisfaction was very high and, in spite of this, the peer review judgment was that most of these contractees were (only) threshold competent. Similarly, 5 out of 6 contractees that had average or higher levels of client satisfaction were assessed as threshold competent by peer reviewers. Unsurprisingly there was no significant correlation between

[91] It focuses on those contractees that can be said with 90% confidence that they performed either significantly better or significantly worse than average. The peer review assessment is made to a 95% confidence interval. A 90% confidence interval is adopted for client satisfaction because of the very low number of contractees who were distinguished as below or above average using a 95% interval, see para. 6.47.

the mean client satisfaction score of a contractee and the mean peer review score.

These results add weight to the view that client satisfaction primarily measures something different from technical quality.[92] Peer reviewers were judging advice and work done from the file, whereas clients were judging cases on the reality of the service delivered as they perceived it. As a result, client satisfaction is an important, but different, aspect of quality from legal and practical expertise. Clients' judgements on service can be misleading indicators of technical quality, as these results and the model client exercise showed: clients (even where they are repeat players in the sense that the model clients were) could be satisfied with confident and helpful service which significantly misrepresented their legal position and undermined their interests.

Peer review and model client visits

Model clients similarly checked on service aspects of legal work but also involved peer reviewers in the assessment of the technical quality of advice. Model client visits involved only one visit to the contractee. Usually, this involved contact with only one adviser from that contractee. For the NFPs this generally involved an assessment of Level I work (possibly carried out by a volunteer adviser, rather than a specialist contracted worker). As a result, the model client visit itself would not be expected to represent a reliable indication of the quality of the whole of a contractee's work. Nevertheless, it did provide important insights into the initial contact between clients and advisers, and given the number of such visits, a reliable indication of the overall quality of contracted work at a key initial interface between the client and the contractee.

As was discussed in Chapter 7, each of these model client visits was assessed by the peer reviewers who provided an overall score on the quality of work conducted for the model client. Some sixteen contractees who were peer reviewed were also visited by model clients. Unsurprisingly, given the one-off nature of these model client visits, the peer reviewer assessment of model client information did not correlate closely with peer review assessment of the contractees across a full range of files although in general the results did not go in opposing directions. The following table provides an indication of the variation.

Table 10.9: Peer review and peer review model clients information compared

Model Client Assessment	Peer Review File Assessment

[92] See, Paterson, A.A. (1990), *Professional Competence in Legal Service* (National Consumer Council) and Sherr, A., Moorhead, R. and Paterson, A. (1994) *The Quality Agenda: Volumes I* (HMSO, London), Chapter 2, p. 10.

	Below threshold competence	Threshold competence	Above threshold competence	Total
Lowest third	1	4	---	5
Middle third	---	5	1	6
Highest third	1	3	1	5
TOTAL	2	12	2	16

This suggests that the one-off model client visits did not provide a significantly similar predictor of the overall legal quality of a contractee's work. It may well be the case that increasing the number of model client visits to assessed contractors would have provided a broader and more consistent view of their overall quality. Each visit was, however, quite labour intensive (for the model clients themselves and those administering them) and extra cost was involved in using peer assessors to review the legal quality of the advice given.

Peer review of model clients, model clients and client satisfaction

Similarly, there was no correlation between client satisfaction and peer reviewer scoring of model client visits. There were, however, significant correlations between client satisfaction and the scores indicated by model clients themselves on three of the model client service questions: whether the model client felt that the adviser understood their problem;[93] whether the adviser gave the impression that they were interested in the problem (as opposed to regarding it as trivial or insignificant);[94] and whether the adviser seemed to deal efficiently with the information that the model client gave and established as complete a picture of the problem as possible.[95]

These assessments were, of course, lay assessments by model clients who had no legal training, although they would have built up some inchoate understandings of quality through repeated visits to different suppliers and the training they received from the researchers. The relationship between the model client findings on service and client surveys suggests that model client type approaches can provide a very useful additional or alternative means of looking at client perspectives (without the problems of low response to which client surveys are prone). The significant correlations suggest that even on small numbers, model client lay-assessments of service could provide an alternative or proxy for client surveys although it would still be necessary to do more than one visit per contractee, with the attendant resource implications. In addition, useful qualitative insights may be added to the lay-assessments of service and technical quality through the involvement of peer reviewers. The latter's involvement is (compared with full peer review) quite cost efficient.

[93] Spearman's RHO correlation coefficient = 0.590, p = 0.003.

[94] Spearman's RHO correlation coefficient = 0.488, p = 0.018.

[95] Spearman's RHO correlation coefficient = 0.459, p = 0.028.

These findings also strengthen an understanding of what it is that client satisfaction is measuring. In chapter 6, a number of factors were identified which appeared to help drive satisfaction, for example: advice on length of case and the use of multiple advisers. The correlation between model client judgments and client satisfaction results suggests that the ability of advisers to convey to the clients that they understand the problem; are interested in it and acknowledge its importance to the client; and seem to deal efficiently with the information the client gives and establish as complete a picture of their problem as possible, may all make lay satisfaction more likely.[96] Most of these factors are not solely client-handling skills, but relate to the advisers' technical competence. Nevertheless, the difference between peer review and client satisfaction and lay model client data has confirmed a crucial difference between lay understandings of quality and peer or technical assessments of quality. The possibility of being effective with clients but ineffective with their legal problems is clearly a reality.

Outcomes, Peer Review and Satisfaction

Chapter 9 discusses outcomes under contract. The following table summarises the position. It was possible to compare a contractee's overall level of positive financial results with their peer review scores in all 52 of the contractees who were peer reviewed.

Table 10.10: Peer review and incidence of positive financial results compared

	PR Assessment (overall)						
Outcomes	Below threshold competence		Threshold competence		Above threshold competence		Total
	n	%	n	%	n	%	
Lowest Third	6	35.3	10	58.8	1	5.9	17
Middle Third	2	11.8	12	70.6	3	17.6	17
Highest third	2	11.1	8	44.4	8	44.4	18
Total	10	19.2	30	57.7	12	23.1	52

As indicated in the table above, 36% of peer reviewed contractees who had levels of positive financial results in the lowest third for all contractees were assessed as below threshold competence of peer review (compared with about 11% of other peer reviewed contractees). Equally, those in the middle third were most likely to be marked as threshold competent, and those in the highest third were most likely to be marked on peer review as at above threshold competent. Thus there is a demonstrable relationship between positive financial results and quality under peer review. The relationship between the proportion of cases in which a contractee got positive financial results and the peer reviewer's assessment of the contractee was statistically significant.[97]

---

[96] A larger number of model client visits conducted in each contractee with parallel client satisfaction surveys of real clients would be needed to be more confident of this.

[97] Pearson correlation coefficient = 0.304, p = 0.028.

That said, the level of positive financial results was not a perfect predictor of poor or good quality.

There was generally no significant correlation between a contractee's average client satisfaction scores and their overall level contract outcomes.[98] In particular, it should be noted there was no significant relationship between clients ceasing to give instructions and levels of client satisfaction measured in the client survey. This suggests that clients ceasing to give instructions is not a strong proxy for low levels of client satisfaction. Similarly, there was no significant relationship between these outcome measures and the peer reviewer's assessment of model client visits.

A correlation was found between positive financial results and client satisfaction on individual cases showing that whilst individual cases with positive financial results had higher levels of satisfaction this had not fed through at a more general level. Contractees that generally got higher levels of results were not shown to get generally higher levels of client satisfaction than other contractees. This may be due to the inability of the client survey to distinguish between contractees in terms of overall levels of satisfaction: most contractees were rated as 'average' on the survey.

This general analysis of outcomes was repeated more specifically within the welfare benefits and housing work categories to look at outcomes within those work categories.[99] Significant (and strong) correlations were found between peer review assessments of contractees on their welfare benefits work and financial results in welfare benefits cases;[100] and also specifically between rates of positive financial results in welfare benefit challenge work and peer review scores in the welfare benefit work category.[101] Similarly in housing work, the proportion of cases in which property was received or retained correlated significantly with peer review scores.[102] In general therefore, peer review seems to correlate with a number of outcome measures but not with the other indicators of quality so far considered (client surveys and model client scores).

---

[98] Although contractee client satisfaction survey rating (using the 90% confidence interval) transposed into 3 statistically significant satisfaction levels (higher, lower or average levels of satisfaction) did correlate with levels of receipt or retention of property. Pearson coefficient 0.264, $p = 0.016$.

[99] These two areas were the ones where there were sufficient peer reviews to enable some judgements to be made on a number of contractees.

[100] Pearson correlation coefficient 0.658, $p = 0.001$.

[101] Pearson correlation coefficient 0.648, $p = 0.001$.

[102] Pearson coefficient 0.426, $p = 0.034$.

Other relationships to quality

Time spent on matters and volume of cases

There was no statistically significant relationship between client satisfaction and the average time spent per matter by contractees on all of their cases.[103] There was not a significant relationship between the volume of matters closed under the contract and client satisfaction. Nor was there any observable correlation between these two variables and peer review scores or peer reviewer assessments of model clients. As a result, on this evidence the amount of time generally spent by a contractee on contracted cases is not likely to act as a useful trigger to prompt use of peer review.

Case profiles

Briefcase data was examined to ascertain whether there was any relationship between peer review assessments of quality and the case profile of contractees. In housing and welfare benefits work, the relationships between proportions of main problems, client types, principal issues and complicating factors were considered to see if they showed any significant relationship to the peer review results.

For welfare benefits work, contractees with higher proportions of contracted cases where first instance court/tribunals, appeals and/or judicial review were complicating factors, tended to have better peer review assessments.[104] Similarly, those contractees that did low proportions of disability type benefits (means tested and non-means tested) were more likely to score poorly on peer review.[105] And those contractees that did lower proportions of welfare benefits work where a refusal of benefit was a principal issue were more likely to score poorly on peer review.[106] This emphasises the importance of welfare benefits contractees being able to deal with some adversarial work.

In housing, those contractees with higher proportions of cases where threatened homelessness was a principal issue tended to have higher peer

---

[103] On a case by case basis such correlations were apparent but they did not hold for contractees general profiles.

[104] The Pearson coefficient score was = 0.485, p = 0.019 between the welfare benefits work, and the proportion of a contractee's caseload which involved first instance tribunals/court hearings, or appeals to tribunals or courts or judicial review. The same was true for cases taken to appeal courts/tribunals in welfare benefits cases.

[105] The Pearson correlation coefficient was = 0.513 , p= 0.012 for the correlation between the proportion of a contractee's caseload where the client's benefit problem related to either means or non-means tested benefits for disability, sickness or injury and the peer review assessment of that contractee's welfare benefits work.

[106] Pearson correlation coefficient = 0.546, p = 0.007. Where the correlation compared cases which involved either a refusal, reduction, or withdrawal of benefit and the peer reviewer's assessment of the contractee's welfare benefits work, the Pearson correlation coefficient was similar (0.512, p = 0.012).

review scores.[107] This similarly emphasises the importance of adversarial work. Very similar results were found for contractees with higher proportions of cases where rent/mortgage arrears was a principal issue.[108] The proportion of cases where the matter involved either judicial review, first instance proceedings, appeals or the matter proceeding on to legal aid was similarly linked to peer review scores.[109] It is possible that the peer reviewers tended to privilege in their assessments those matters which had a higher degree of court/tribunal exposure or were more serious. Dealing with such cases would need a higher level of technical competence and confidence. Their approach may well be appropriate.

The next section considers the extent to which such trends would aid a prediction of poor performance (as measured by peer review) and so help the Commission to select contractees for peer review, or in general terms determine quality in accordance with the approach of peer reviewers.

Can outcomes and case profiles help target review?

The following table summarises the three main factors that, as indicated above, significantly correlated with peer review assessments of quality in welfare benefits work.[110] The table exhibits all contractees that had their welfare benefits work assessed by peer review, and ranks contractees by their peer review score (right hand columns of the table). The proportion of welfare benefits work and disability-type benefits is indicated against each contractee, as is the proportion of cases which involved first instance and appeal work with tribunals and/or courts and judicial review as complicating factors. Under each of the column headings, a + sign indicates that the value was significantly higher than average, and a – sign indicates that the value was significantly lower than average. Where there is no sign, the contractee's profile was not significantly different from the average.[111]

Table 10.11: Welfare benefits quality triggers

Positive Financial Results	Welfare Benefit Challenges	Court, tribunal and judicial review proceedings (incl. Appeals)	Disability-type benefit clients	Peer review assessment of welfare benefit's work

[107] Pearson correlation coefficient 0.432, p = 0.031.

[108] Pearson correlation coefficient 0.419, p = 0.037.

[109] Pearson correlation coefficient 0.479, p = 0.016.

[110] Welfare challenges, and the extent to which principal issues included refusals, withdrawal or reduction of benefits were strongly related to each other (contractees with high proportions of challenges also tended to have high proportions of refusals, etc., and vice versa. As a result, for simplicity, welfare challenges were chosen as the predictor here.

[111] For example, some of the peer review scores were not significantly different from the mean because of the lower number of files looked at and/or the variation in the marks that they achieved on files.

Contractor		Position from average		Position from average		Position from average		Position from average		Position from average
10	63%	+	74%	+	36%	+	72%	+	3	+
21	28%	+	74%	+	24%	+	56%	+	3	+
141	24%		30%	-	0%	-	16%	-	3	+
6	41%	+	57%	+	15%	+	65%	+	2.9	+
32	63%	+	100%	+	69%	+	69%	+	2.7	+
50	35%		81%	+	51%	+	48%	+	2.5	
60	3%	-	20%	-	7%	-	38%	+	2.4	
45	8%	-	61%	+	15%	+	36%		2.4	
19	58%	+	52%	+	26%	+	56%	+	2.2	
7	5%	-	20%	-	2%	-	26%	-	2.2	
37	11%	-	45%	+	6%	-	32%		2	
64	8%	-	18%	-	2%	-	23%	-	1.9	
56	17%		43%		7%		14%	-	1.8	
71	27%		71%	+	24%		0%	-	1.7	
110	0%	-	8%	-	0%	-	2%	-	1.7	
53	4%	-	27%	-	1%	-	19%	-	1.6	-
117	0%	-	60%	+	20%		12%	-	1.6	
83	1%	-	63%	+	31%	+	29%		1.6	
77	0%	-	33%		0%	-	17%		1.5	
16	27%		80%	+	0%	-	68%	+	1.5	-
67	9%	-	19%	-	0%	-	39%	+	1.3	-
58	4%	-	18%	-	1%	-	31%	-	1.2	-
124	0%	-	10%	-	0%	-	50%	+	1	-

5 out of 22 contractees had their welfare benefits work assessed at significantly below threshold competence (23%). The table shows that of the 13 contractees that had below average levels of positive financial results in welfare benefits challenges, 9 (69%) were assessed at below threshold competence of which 4 (31%) were *significantly* below threshold competence. Similarly, of the 9 contractees who did less than the average proportion of welfare benefits challenges, 6 out of 9 (67%) were assessed at below threshold competence and 4 contractees (44%) were performing *significantly* below threshold competence as assessed by peer review. Of the 12 contractees who did significantly less than average amounts of court, tribunal and judicial review related work, 8 out of 12 (67%) were assessed at below threshold competence and 5 were assessed as *significantly* below threshold competence on peer review (42%). 7 out of 9 (78%) of contractees with lower than average levels of disability benefits work were assessed as below threshold competence, although only 2 of these (22%) were significantly below threshold competence.

These figures suggest that by targeting a significant proportion of peer reviews at a sample of contractees that carried out below average proportions of welfare benefits challenges; advice and assistance relating to welfare benefits

tribunals and associated court, and judicial review proceedings; and/or fewer disability-type benefit clients the effectiveness of peer review in identifying poor performers could be increased. Similarly, targeting peer review at contractees with lower levels of positive outcomes will improve the likelihood of peer review identifying contractees that are performing poorly. Conversely, the use of such triggers is not a perfect guide to the quality of an organisation's work and could not be used as such;[112] rather it provides a means to target (say) peer review at a group of contractees most likely to give rise to quality concerns. The 'Hawthorne effect' also needs to be borne in mind if the Commission were to adopt outcome and profile measures as indicators of quality concern: contractees aware of the importance of certain case profiles and outcomes would be likely to adapt their behaviour and recording of contracted matters to meet outcome and profile targets. A strategy for dealing with these problems and developing and refining quality triggers is set out below.

[112] Contractee 141 scored poorly on the outcome and case profile indicators but had amongst the best ratings on peer review, whereas contractee 16 scored highly on 2 out of 4 indicators but poorly on peer review.

The next table performed a similar comparison for housing work.

Table 10.12: Housing quality triggers

	Received/retained property outcomes and/or Third Party Action		Threatened homelessness (a principal issue)		Tribunals, courts, judicial review or legal aid (a complicating factor)		Peer Review Assessment in Housing	
Contractor	Proportion of caseload	Position from average	Proportion of caseload	Position from average	Proportion of caseload	Position from average	Mean score	Position from average
22	80.2%	+	98.8%	+	96.5%	+	3.0	+
5	29.6%		42.7%	+	29.8%	+	2.8	+
85	56.7%	+	50.9%	+	19.8%	+	2.8	+
19	50.0%	+	48.0%	+	41.0%	+	2.7	+
127	21.1%		48.1%	+	11.5%		2.6	+
80	42.0%	+	18.3%	-	9.8%	-	2.4	
104	27.5%	-	41.6%	+	36.8%	+	2.4	
83	20.7%		30.2%		13.2%		2.3	
30	31.7%		30.7%		9.6%	-	2.2	
50	18.5%	-	28.6%		28.6%	+	2.2	
100	0.0%	-	0.4%	-	5.3%	-	2.1	
15	53.8%	+	35.4%		6.5%	-	2.0*	
51	17.0%	-	41.0%		17.9%		2.0*	
89	31.9%		33.5%		5.7%	-	1.9	
91	26.9%		34.6%		17.3%		1.9	
143	18.0%	-	28.0%		37.7%	+	1.9	
21	27.4%		29.5%		17.0%		1.8	
94	26.3%		10.0%	-	2.5%	-	1.8	
126	32.5%		30.7%		9.6%	-	1.8	
135	25.8%		27.3%		3.0%	-	1.8	
71	23.8%		37.4%		22.0%	+	1.5	
23	36.4%	+	48.5%	+	20.1%	+	1.4	-
47	40.7%	+	42.6%		18.5%		1.3	-
115	26.0%		36.0%		18.0%		1.2	-
67	12.4%	-	18.6%	-	6.2%	-	1.0	-

*Actual values = 1.95, rounded up to 2.0 to one decimal place.

14 out of 25 contractees had average peer review scores of below threshold competence (56%) on their housing files and 4 out of 25 (16%) contractees were rated as significantly below threshold competence. Of the six contractees who had below average proportions of cases where property was received/retained and/or beneficial third party action was taken three (50%) had peer reviews at below threshold competence but only one (17%) was significantly below threshold competence. In other words monitoring contractees on this did not help predict whether they would be more likely to be assessed at below threshold competence on a peer review. Of the four contractees with below average levels of threatened homelessness as a principal issue, two (50%) were below threshold competence, one (25%) significantly so. Of the nine contractees with below average levels of tribunal, court, judicial review and legal aid matters six (75%) were below threshold competence; although only one of these was significantly below threshold competence (11%). This combined factor may provide a more useful indicator on which to base targeting a sample of peer reviews, but the results in housing suggest that triggers for peer review need careful ongoing development.

In addition to the data collected by the research, data relating to the Commission's audits was also considered. This is discussed in the next section.

Management audits by the Commission

During the life of the pilot, data was requested on the management audits conducted on contractees by the Commission. This information was incomplete and fresh requests were made for data. The intention was to compare this data with peer review assessments of contractees. Data on 39 of 52 peer reviewed contractees was made available, but concerns remained about its completeness.

Three sets of data were compared. The number of major non-compliances recorded during management audits was compared with the peer review assessment of contractees, as were the last available transaction criteria audits on that contractee in each relevant work category.[113] There were no discernable relationships between the number of major non-compliances and the peer reviewer assessment of contractees work. Nor were there any discernable relationships between the number of audit observations in the most recent contract audits.[114]

A more qualitative assessment of the nature of management audit failures for contractees that had failed peer review assessments suggested there may be a linkage between supervision and file review requirements and the peer reviewers' assessment of quality. The Commission's management audit data was available on 6 out of the 11 contractees that had failed peer review. In relation to five of these (83%), there were concerns expressed about the contractees' ability to meet supervisor requirements, to supervise the work and/or the conduct of file reviews in at least one audit since 1997. A reading of other audit reports (of firms that either passed peer review or were not reviewed) suggested that about one half of contractees have had this concern raised in recent audits (probably reflecting the increased emphasis on these issues as a result of LAFQAS). Thus the fact that audits had revealed some concerns relating to supervision and file review will not by itself enable the Commission to be confident that poor performing contractees should be identified or targeted for peer review on this basis.

A comparison between the Commission's transaction criteria audits and peer view scores was also inconclusive, because of a lack of recent transaction

---

[113] Comparisons were against transaction criteria audit scores from the last audit during the life of the pilot (or the nearest transaction criteria audit in each relevant work category, if no such audits were conducted during the life of the pilot).

[114] These were very few in number and so it may be too early to assess any linkage between peer review and this aspect of contract audits which, rather than focusing on failures to meet the management standards, identified more subjective areas where in the auditor's view there was room for improvement.

criteria data on contractees that had been peer reviewed.[115] There were significant correlations between transaction criteria scores and peer review scores in debt cases and housing but not in welfare benefits. All of these assessments were based on too small a sample to draw any reliable conclusions in any event. This comparison should be repeated under research conditions when more data is available.

Size of contracts

Another area which was investigated was whether the size of contract had any consistent effect on the level of quality in a contractee (as measured by peer review). No such relationship was found.

Conclusion: A strategy for using outcome and/or case profile measures as quality triggers

The results for welfare benefits peer reviews and, to a lesser extent, the housing peer reviews, suggested the utility of monitoring case profiles and outcome results as an 'early-warning' trigger leading to peer review in the event that a contractee started to slip on some of the indicators.

One likely effect of this approach is that formally, or informally, contractees will learn of the triggers and start to adapt their recording behaviour, and possibly the type of cases they take on and how they handle them, to meet profile and outcome triggers (which would most likely be understood as 'targets'). The utility of such triggers might diminish as a result of this experiment effect. A second observation is that such triggers in any event need adapting and refining on a larger number of contractees and in the light of any 'experimenter effect'. A third observation is that, as far as is possible, any triggers/indicators should be desirable in themselves: therefore indicators which look at case profiles would need to weigh the ability to predict poor quality with the desire to encourage balanced case profiles in contractees.

If the Commission is minded to introduce peer review, it would be sensible to adopt a two-track approach to targeting reviews. In particular, throughout any peer review programme it would always be important to ensure that at least part of the peer review sample was targeted on an entirely random basis at contractees. This would provide some protection against contractors who score well on paper indicators from evading peer review. (See for example, contractee 16 in the Welfare Benefits Table 10.4). It would also provide a means for continuing to monitor and to define any peer review triggers (see below). Equally and quite importantly, it would also ensure that peer reviewers were sensitized to the full range of quality in contractees and would not simply be exposed to contractees who are expected to perform poorly.

---

[115] 15 housing contractees had been peer reviewed and had some transaction criteria audit data to enable a comparison. This was true of 9 welfare benefits contractees, 2 debt contractees and 1 employment contractee.

The second sample of peer review visits could be targeted at those who, because of outcome and case profiles, might be expected to perform more poorly on peer reviews. Welfare benefits work (see above) seems particularly susceptible to this approach (e.g. having low levels of positive financial outcomes acting as a trigger). It should be possible through the random selection of peer reviews and the continuing development of peer review to develop further, or refined indicators.

There is potential for triggers to include the outcome of management, contract compliance and transaction criteria audits. Results from the model client visits suggest that a particular failing that audits could look for as a trigger is the failure to provide the client with practical advice on the steps which the client or an adviser can take to remedy or mitigate the client's problem. Similarly, peer review could be used as part of the process for improving the appropriateness and approach of the Commission's routine audits. However, it is too early to say whether there is a significant link between peer review and the Commission's management data and transaction criteria audit data. It is conceivable that the Commission could develop a more subjective aspect to ordinary contract audits which could, in itself, act as a trigger to peer review.[116] This is a process that would need to be developed carefully and openly. Whilst management reviews were carried out on an organisational basis, transaction criteria review is often carried out on a very small number of files. Widening the area of transaction criteria review, prior to organising a peer review may be a sensible and efficient approach. Building information on experience of both will help to provide more certainty in the assessment of quality.

Peer review has the potential to significantly improve the Commission's ability to identify poor performers and take the appropriate action. It can also be used to develop and test existing quality indicators and assurance mechanisms. It also has the potential to provide constructive feedback to contractees. To meet the last aim, peer review needs sensitive and careful implementation.

Chapter 11 brings together findings on quality as a result of this research and makes further recommendations as to the utility of quality measures. This chapter has sought to look, in particular, at a possible method for building peer review into the assessment of contracted work and more generally into the quality assurance systems of the Commission. Peer review has a strong role to play but it is one which needs to be carefully managed. The development of triggers needs to be part of an ongoing rigorous process of assessment and development.

---

[116] The Commission has recently introduced as part of its audit process a "summary report" which assesses in broad brush terms: a) how easy the file was to audit; b) how well client care was handled on the file; and c) whether advice given was satisfactory. This research project has not been able to assess the utility of the Commission's approach to this, given its recent introduction.