

THINKING SKILLS

Paper 9694/21
Critical Thinking

Key Messages

The main reason why candidates lose marks seems to be that they do not understand what questions expect them to do. On this occasion, this applied especially to **Question 3(c)**. Candidates and teachers are strongly advised to study previous question papers and mark schemes in order to find out what kind of answers are expected in response to the different types of question.

Candidates and teachers need to understand that disagreeing with claims or evidence does not constitute an evaluation of reasoning.

There are not many technical terms in Critical Thinking, but candidates are expected to understand and be able to use the specialised meanings of “argument” and “assumption”.

General Comments

There were a few good scripts, but most marks fell within the lower half of the range. Some candidates omitted part of the exam, while a few appeared to run out of time. Some of the latter wisely chose to answer **Question 3(d)** before **3(c)**.

Comments on Specific Questions

Question 1

Almost all candidates seemed able to understand both the scenario and the questions.

- (a) A few candidates correctly explained that being located outside the country enabled the paper to say what it wanted, without being influenced by the government, while the inability to see at first hand meant that the reports might be inaccurate. Many of the answers were vague, and some did little more than restate the information given in the question: these candidates might perhaps have been thinking on the right lines, but their answers could not be credited, because they did not actually make the required points explicitly.
- (b) This question was fairly straightforward to candidates who knew the technical meaning of the expression “unstated assumption”, and most of those achieved 3 marks out of 3. But many candidates answered the question by quoting from Source C. A quotation is not “unstated”, and therefore can never be the answer to a question of this type.
- (c) A fair number of candidates correctly identified the most significant alternative explanation (that the soldiers were present in order to intimidate voters into voting for the President) and were awarded 2 marks out of 2. Other reasonable answers were awarded 1, while the few answers which were entirely speculative were judged to deserve 0 marks.
- (d) Opinions were divided roughly equally between the two main options, namely that the President won the election fairly or unfairly. Very few candidates went for the middle option, that he cheated but would have won anyway. Quite a lot of good answers were offered, and a fair number of candidates were awarded 5 or 6 marks. A significant number of candidates addressed issues of credibility of sources and/or used inferential reasoning. The weaker answers tended simply to believe or reject the allegations contained in Source C, and judge accordingly. Many candidates assumed that because (according to Source A) Joe Vos had been accused of corruption, he must be guilty. Although a lot of candidates commented that the credibility of the international observers

(Source E) was enhanced by their neutrality, very few pointed out that their ability to see and to understand was probably limited.

Question 2

Not many candidates had the confidence to recognise that these sources were of very little value.

- (a) Most candidates achieved 4 marks out of 4 on this question, by recognising that the support given for the claim was quite weak. The minority of candidates who identified only one side of the evidence achieved a maximum of 2 marks.
- (b) The focus of this question was the representativeness of evidence. Only a minority of candidates focused on the contrast between home and hotel, and relatively few realised that couples are unlikely to argue about the side of the bed at home, because they made the choice when they moved in.
- (c) One of the aims of a course in Critical Thinking is to encourage candidates (and the adults they become) to be suspicious of the motivation behind commercial claims. This question was intended to help candidates see that the research was undertaken in order to provide the subject matter for a news story, which would be supplied to and published by newspapers and magazines, with the intention of advertising the hotel chain and the improvements it had recently made to its rooms. Most candidates, however, took the claims at face value, and stated that the motivation of the research was to give the best experience to clients.
- (d) Nearly all candidates achieved some marks on this question, by correctly drawing relevant inferences from one or more of the sources. The best answers rejected the claim, recognising that the support offered by the sources was distinctly weak, since Source A did not really establish what its author thought it did, while the other two sources were advertising ploys rather than serious research.

Question 3

- (a) Not many candidates correctly identified the main conclusion of the argument. The most popular answers were the last sentence of paragraph 2, presumably because it included the word “therefore”, and all or part of the final paragraph.
- (b) As usual, the passage consisted of several mini-arguments, each of which led to an intermediate conclusion, which in turn supported the main conclusion. These ICs were the correct answers to this question, but the most popular answers were the starting-points of these mini-arguments, instead of their conclusions.
- (c) As in previous sessions, the most popular approach to this question was to disagree with the author’s argument, either as a whole or point by point, instead of evaluating its reasoning. Answers of this kind were awarded 0 marks. Even many of those who did spot weaknesses in the reasoning expressed them as challenges instead of using more precise or technical language, and thereby scored 1 mark instead of 2 for each point. Some candidates identified strengths in the reasoning, which were not credited. Some people must have wasted a significant amount of time writing long answers which achieved 0 marks. Study of the mark scheme will reveal the valid evaluative points which were credited and should give some indication of what to look for in future series. A number of candidates challenged the relevance of the reference to Boethius in paragraph 3, while some rejected the claim that babies are rational (although few, if any, correctly identified this flaw as equivocation), and some rightly criticised the generalisation in paragraph 4.

As in previous series (and in **Question 1(b)**), nearly all candidates who referred to “unstated assumptions” wrongly interpreted that expression as meaning “unsupported statements”. It is never correct to claim that a quotation from the passage is an unstated assumption, because if it can be quoted, it is not unstated.

- (d) As usual, a complete range of marks was awarded, with many candidates achieving 3 marks and a good number 4 or 5. Quite a lot of candidates limited their mark to 3 out of 5 by discussing the value of parenting, without mentioning “improved” or making a connection with “a better world”. A few candidates discussed the ideas from the passage instead of producing an independent argument supporting or challenging the claim supplied.

THINKING SKILLS

Paper 9694/22
Critical Thinking

Key Messages

- Little credit can be given for answer content which merely repeats what is in the passage. Many candidates waste time by doing this and typically gain any marks for one or two sentences at the end of their answer. This is often the reason why the answers to 3-mark questions are over-long. This was particularly noticeable in **Question 1(a)** in this paper.
- There seemed to be an increase in the number of candidates who dealt with **Question 3(d)** by treating it as an essay question looking at both sides of the issue. Counter-arguments must be clearly replied to in order to get credit as contributing to the overall thrust of the reasoning. Candidates cannot obtain full marks if they do not sustain one clear argument *either* for *or* against the proposition offered.

General Comments

The examination produced a good range of marks and was, overall, consistent with previous papers as regards style and difficulty of questions. Rather poor performance in **Question 3(c)** was compensated for by good performance in **Question 2(d)**. Candidates seemed to respond well to the issues raised by the questions and were able to tackle them effectively. As in previous series, some candidates need to understand that expressing opinions about the issues raised or showing further knowledge of them is not the focus of the paper and cannot receive much credit, if any. A significant minority of candidates spend too much time on **Question 1**, meaning subsequent questions are rushed. In particular, they failed to get on to **Question 3(d)** where the marks are more accessible than in other questions. Such candidates also tend to give over-long answers on the 3-mark questions in **Questions 1 and 2**.

Comments on Specific Questions

Question 1

- (a) This was not done well, with few candidates seeing that an assessment of the *overall* effects depended on the evidence for long-term effects. While such evidence came from a government report, it only showed a *lack* of evidence rather than definite evidence there were no long-term effects. Some candidates managed to gain 1 or 2 marks by suggesting or implying that if the effects were only short-term then this diminished the significance of the evidence.
- (b) Very few candidates considered both sides of the question of the reliability of the evidence. Most opted for the point that this was unreliable media hype without considering how the points made were corroborated by expert evidence in Sources D and F, thus increasing the reliability of the evidence.
- (c) This question was done well, with many candidates getting 3 marks and seeing the problems with moving beyond the suspicions raised by this case to a generalised conclusion and one which showed a definite causal link.
- (d) A large number of candidates found it difficult to move beyond Sources B and C, but they were able to make evaluative points and inferences, and almost always considered an alternative conclusion. Candidates should remember that the alternative conclusion must be plausible – it does not seem plausible that the water company planned to poison their customers, as one or two candidates suggested.

Question 2

- (a) This question discriminated well, with good answers noting that it did not enable this conclusion to be drawn, usually on the grounds that we had no information on the different rates between countries where neonicotinoids are used or not used. Less good answers discussed the credibility of the source making this claim, which was not the focus of this question.
- (b) A majority of candidates correctly identified this as an argument, though their explanations were sometimes not convincing enough to gain a mark (e.g. '*It is an argument because it has evidence*'). Convincing explanations were usually helped by a correct identification of the main conclusion.
- (c) This was a difficult question, but a sizeable minority of candidates managed to correctly identify the two key bits of information: whether neonicotinoids contained copper and whether they were used in conjunction with other pesticides.
- (d) This question was done well, with a large number of Level 2 and 3 answers. Candidates used the sources effectively to support their conclusion, often making correct inferences and evaluations. A number correctly saw that much of the evidence about harmful effects was only relevant to neonicotinoids. Others explored the difficulty in isolating pesticides as a causal factor from all the other possible causes.

Question 3

- (a) Most candidates gained 1 mark, but only a minority correctly identified the main conclusion as opposed to the tempting reason that led to it. A minority were misled by the 'So' in the last sentence of paragraph 4 into thinking that the main conclusion was 'The distinction between the technology and the way people use it is meaningless,' in spite of the fact that this had little relation to the overall thrust of the argument in the passage.
- (b) In this question also, most candidates gained 1 or 2 marks and a significant number gained 3. Having identified one of the reasons as the conclusion in part (a) did not seem to unduly hamper candidates in identifying other reasons.
- (c) There has been an increase in the number of candidates who understand that this question is about evaluating reasoning rather than challenging propositions, but many candidates found it difficult to successfully make evaluative points. Many referred to *ad hominem* and *straw man* but these flaws did not figure in this passage of argument. There was a tendency for candidates to get rather bogged down in the content of paragraph 4. They correctly saw that the comparison with the car was not a good one but were unable to explain this clearly. They also tended to challenge the statement that congestion and pollution would disappear if nobody drove cars. Paragraph 5 also tempted candidates in to challenging the statements made rather than the reasoning used. The overall result of these difficulties was that 4 or 5 mark answers were uncommon.
- (d) Candidates used the proposition effectively and there were many 4 or 5 marks answers, but many candidates spoil an argument that was developing well by suddenly lurching in the opposite direction (as mentioned in the Key Messages). Given the usual assumptions about young people and technology, a surprising number argued in favour of the proposition. Very few candidates challenged the extreme nature of the statement by seeing the distinction between 'humanity being destroyed' as opposed to 'human life moving in an undesirable direction': even good arguments in favour of the proposition tended to conflate the two. A minority of candidates did successfully argue for 'humanity being destroyed', as such, via the idea of intelligent machines taking over the world.

THINKING SKILLS

Paper 9694/23
Critical Thinking

Key Messages

The main reason why candidates lose marks seems to be that they do not understand what questions expect them to do. Candidates and teachers are strongly advised to study previous question papers and mark schemes in order to find out what kind of answers are expected in response to the different types of question.

Candidates and teachers need to understand that disagreeing with claims or evidence does not constitute an evaluation of reasoning.

There are not many technical terms in Critical Thinking, but candidates are expected to understand and be able to use the specialised meanings of “argument” and “assumption”.

General Comments

There were some very good scripts, and very few if any which did not show some evidence of Thinking Skills. Some candidates omitted part of the exam, while a few appeared to run out of time. Some of the latter wisely chose to answer **Question 3(d)** before **3(c)**.

Comments on Specific Questions

Question 1

Almost all candidates seemed able to understand both the scenario and the questions.

- (a) Most candidates scored at least 2 marks out of 3 for this question, although many of the answers were judged to be undeveloped or marginal. The most significant answers focused on ability to see and expertise. A popular marginal answer was that the security officer was unbiased, but the fact that the reader does not know that the officer is biased is not the same as knowing that he is not. Some candidates claimed that the security officer had many years' experience, but there is no basis for this claim in the source. Candidates from a few Centres seemed not to know the kind of answers which were expected to a question about reliability.
- (b) Most candidates scored 3 out of 3 for this question, by identifying two valid criticisms of Dr Cameron's evidence and developing at least one of them.
- (c) A lot of candidates scored 0 for this question because they summarised the content of the source without drawing any inference from it. Very few suggested that Mrs Breslau was probably a kleptomaniac.
- (d) A lot of good answers to this question were offered, and a good number of candidates were awarded 5 or 6 marks. Many of them addressed issues of credibility of sources and/or used inferential reasoning. The most popular and most plausible answer, on the basis of Source D, was that Mrs Breslau had stolen the shirts. Nearly everyone proposed rational reasons for the theft (that the shirts were for her own use or to give away or sell), rather than suggesting that she was a kleptomaniac. Some candidates declined to come to a decision about how the shirts had found their way into Dr Breslau's bag, but this is not a good strategy, since the question required them to come to a conclusion of this kind. Solutions which were entirely speculative and unsupported (such as that a personal enemy had “framed” Dr Breslau) were not credited. A very few candidates scored 0, because they neither made use of the sources nor came to a rational conclusion.

Question 2

Very few candidates made the general point, relevant to most parts of this question, that if someone dies while receiving medication for an illness, it is more likely to be the illness which causes the death, rather than the treatment.

- (a) A few candidates recognised that Source A did contain an argument, in which the conclusion was the imperative, “contact us now”. The many who scored 0 marks were divided between those who did understand the meaning of the technical term “argument”, but made a wrong judgment, and those who appeared not to understand the specialised meaning of the word.
- (b) Most candidates scored 3 out of 3 for this question, by identifying two valid weaknesses in the evidence and developing at least one of them. A few received a reduced mark because they paraphrased the facts about Charles Crane’s lifestyle, illnesses and medication without explaining how they weakened the support for the claim.
- (c) Most candidates correctly identified expertise and vested interest respectively as the key factors strengthening and weakening Professor Drek’s evidence, but many scored only 1 out of 2 in each case, by not explaining the points, e.g. by not stating the area of his expertise (the fact that he was an expert was stated in the heading for his evidence) and not stating what he had a vested interest to do.
- (d) Nearly all candidates achieved some marks on this question, by correctly drawing relevant inferences from one or more of the sources. Some excellent answers were received. The best answers judged that Bradiloxx did not cause heart disease, although the opposite conclusion was also credited. Some candidates over-stated the inference that could be drawn from sources, claiming for example that Source B “proved” that the drug was the cause of Senator Blodz’s death. Many candidates misunderstood the nature of Source A, thinking it came from the manufacturers of Bradiloxx rather than a firm of lawyers. The best answers recognised that this source gives very little support to the claim that Bradiloxx causes heart disease.

Question 3

Many candidates found it hard to keep a cool head in analysing and evaluating this argument, because they had so little sympathy with its sentiments.

- (a) Most candidates correctly identified the main conclusion of the argument, even though it occupied an unfamiliar and not obvious position in the passage. The most popular wrong answer was the last sentence of paragraph 3.
- (b) As usual, the passage consisted of several mini-arguments, each of which led to an intermediate conclusion, which in turn supported the main conclusion. These ICs were the correct answers, but the most popular answers were the starting-points of these mini-arguments, instead of their conclusions. Very few candidates performed well on this question.
- (c) More candidates than in previous series approached this question in the right way, but quite a lot of them still scored 0. As usual, some argued against the passage, either as a whole or point by point, and scored 0 despite having apparently put a lot of thought into their answers. Even many of those who did spot weaknesses in the reasoning expressed them as challenges instead of using more precise or technical language, and thereby scored 1 mark instead of 2 for each point. Quite a lot of candidates spotted the exaggeration at the end of paragraph 1, but few if any commented that this did not seriously weaken the point being made. A number of candidates criticised the use of the word “some” in paragraph 3: it was valid to criticise the unjustified move from the “some” in the second sentence to the “all” implied in the final sentence, and this was credited, but the author neither claims nor assumes that all police recruits are motivated by the opportunities for corruption, and the allegation that he does make this claim was not credited. Study of the markscheme will reveal the valid evaluative points which were credited and should give some indication of what to look for in future series.

As in previous series, nearly all candidates who referred to “unstated assumptions” wrongly interpreted that expression as meaning “unsupported statements”. It is never correct to claim that a quotation from the passage is an unstated assumption, because if it can be quoted, it is not unstated.

- (d) As usual, a complete range of marks was awarded, with many candidates achieving 3 marks and a good number 4 or 5. Most candidates chose to support the claim, but there was also more than one acceptable way of challenging it. Some of the reasons in support were rather trivial. Some candidates scored 1 or 0 because they explained how the claim might work out in practice, or the exceptions to it, instead of reasoning in support of (or challenge to) the claim. A few candidates apparently thought that they were supposed to relate the claim to the police, since that was the subject of the passage to which parts (a), (b) and (c) related.

THINKING SKILLS

<p>Paper 9694/31 Problem Analysis and Solution</p>
--

Key Messages

The presentation of orderly working is a decisive factor in gaining marks when a candidate's solution is not correct. A systematic approach, which shows awareness of an audience, is a common feature of most of the high-scoring scripts.

Candidates should consider what evidence can be offered when answering questions which require a judgment or comment – a precise appeal to the relevant data is almost always required for full marks on such questions.

General Comments

Most candidates attempted all four questions.

The questions required no modelling of shape and space, but depended fairly heavily on systematic listing, as a means to unravel the intricacies of their internal logic.

Candidates found **Questions 1 and 2** easier than **3 and 4** in general. **Question 4** required the least logical 'insight', although the performance on it was relatively poor.

Comments on Specific Questions

Question 1

This question required candidates to investigate the interactions of three players in a simple game, which could be modelled using a probability space. Those who employed a systematic listing of outcomes tended to perform strongly on this question.

Fractions were not required for the probabilities, although they were the natural medium for considering the probability questions. Candidates who offered their probabilities in terms of odds (e.g. '3 to 1') rarely managed to answer correctly and unambiguously, and it is strongly recommended that candidates avoid this in future aleatoric Problem-Solving exam questions.

- (a) This question was correctly tackled by only a few candidates – many falling into the trap of oversimplifying the probability space (for example, considering only four options, HHH, HHT, TTH and TTT). It is certainly important that candidates develop an appropriate wariness of probability problems, particularly with regard to the importance of considering whether the order of events is important. With such wariness comes the use of appropriate modelling techniques.
- (b)(i) This question reduced the probability space to a pair of players choosing between 3 choices, and proved more accessible to candidates. The order of events still matter in modelling the equally likely outcomes, but the limitation to two successive events made the logistics easier to 'see'. Some candidates offered verbal descriptions of the probability ("equally likely for all players") which were credited – although it is recommended that candidates conceive all probabilities as fractions if possible.
- (ii) Most candidates were able to see/deduce that it was impossible for Luke to win with a 1.

- (c) (i) This question involved a simple choice of tactics for John, which probably fitted with an intuitive answer to the question. Most candidates completed this successfully. The probability of success was limited to a coin toss and most candidates managed this successfully.
- (ii) This provoked quite a few erroneous responses – some candidates appearing to identify John’s probability of failure ($1 - 0.5$) with Luke’s probability of success. The fact that the best Luke could achieve was to equal John’s score (and still forfeit the ice cream) was appreciated by about half the candidates.
- (iii) This question required a fresh look at the options available when two players are competing, each with two equally likely outcomes – without a means of representing the outcomes, many candidates constructed their proportions clumsily and were not able to reach a correct answer.
- (d) This question required candidates to describe a strategy which required no communication between the players after the games had begun, and which ensured they won equally often. About half the candidates managed to do this. Those who did not tended to offer a system which prevented Mark from winning, but did not ensure that the other won half the time (e.g. Luke plays 3 and John plays 1), or did not explicitly articulate how the two players would ensure they won only half the time – answers which involved the two players flipping a coin to decide who played high and who played low were credited.
- (e) This question did not elicit many correct responses. Many misinterpreted the question as merely asking whether it was possible that Luke gets the ice-cream more than half the time (to which the answer was “yes”, given the fluctuations of probability); others underestimated the task of justifying why not – simply pointing out that Luke could not expect such winnings if John played ‘3’ every time. A correct answer had to appeal to the symmetry of the player’s strategic choices, or explicitly argue that whatever Luke does, there will be a strategy that John could play which would win at least half the time.

Question 2

This problem involved a fairly simple operation (adding and subtracting slices of bread, daily) being iterated over a period time, and required candidates to identify patterns and necessary conditions that could act within the process. The size of the numbers involved introduced an element of arithmetic reliability that challenged some candidates – it also prevented the internal logic from being so ‘visible’ to candidates, as the problem built up.

- (a) (i) Most candidates managed to perform the appropriate calculations for three successive days, and reach the appropriate answer here: $(100 - 88) + (80 - 88) + (100 - 88)$.
- (ii) The process, iterated to the fifth day, led to 0 slices being left. Most candidates managed this without a problem.
- (iii) It was intended that candidates step back from the data, reflect upon their answer to (a)(ii), and then spot the cyclical pattern in the loaf numbers in order to answer this question. Many appeared to do this, and some simply listed the left-overs for the entire 3 weeks: this is a good example of why problem-solving is best tackled in two separate mental modes – one of intense unwavering calculation/processing, and one of probing creative search for efficient strategies.
- (b) (i) This question invited one of three possible answers (more/less/same), but it expected some justification – as stated on the front of the paper, candidates are expected to show their working, and marks can be lost if this is not done adequately. Sufficient support for the judgement (“less”), extended to a clear statement of the difference in loaves between the two schedules, or more detailed working. Many candidates were not awarded the marks here, as a result of insufficient justification.
- (ii) This question required careful processing (to calculate the number of slices eaten according to the second table) and about half the candidates managed this correctly. This question was a test of precision processing, and it may be worth addressing this explicitly in class if many candidates are not using their calculators with sufficient care.

- (c) A comparison of the patterns of ordering shows that Fred's plan would never have ordered 5 loaves on a later date than actually happened, so there would not have been a shortage. Around half the candidates came to this conclusion and scored the mark.
- (d)(i) This question required candidates to consider the conditions that prompted additional loaves being ordered, and then scrutinise the data for the lowest number of slices left over which was followed by an order of 4 loaves. A comprehensive analysis of this is shown in the mark scheme – and yields the answer 11. About a quarter of the candidates managed to make the correct inference here.
- (ii) This final level of abstraction in the question required candidates to consider the mechanism inferred in (d)(i) and compare it with the data, to support a conclusion on whether 'similar usage' could lead to a shortage. There was some scope for sensible interpretation of 'similar usage', but it had to be rooted in the data – and some consideration of what was the highest number seen so far. This needed to be explicitly seen, in order for a judgment to be supported. The triggering mechanism which candidates explored in (d)(i) could then be applied to reach a conclusion. Some candidates took 'similar usage' to mean 'exactly these ordered data repeating indefinitely' and gained no marks.

Question 3

This question required candidates to model the evolution of a game, bound by a small collection of tightly wrought rules. Each round of the game involved consideration of the number of tokens, which decided the price – according to a fairly simple linear rule; and then consideration of a maximum allocation, determined by the number of tokens requested. There was one example given for this latter process, to allow candidates to check that their understanding was correct. It was obviously vital to establish this before launching into the investigative tasks that the question required.

- (a) This question required candidates to continue the sequence 99, 98, 97, ... down to 50, and then deduct 0.5s. A surprisingly small number of candidates managed to do this correctly; the most common error was to suppose that it took 49 tokens to reach \$50, and then subtract ($76 \times 0.5 =$) 38 which yielded the answer \$12. This mistake is easy to make – and candidates' attention should be drawn to it, as a classic error requiring special caution.
- (b) Most candidates managed this process correctly – appreciating that the maximum allocation could be 31, given that they had all over-bid.
- (c) Understanding of the 'maximum allocation' process was tested in this question – and about half of those who attempted the question managed it correctly. Initially, the setting of the maximum allocation may have been done by trial and improvement – although it could be reached by repeatedly subtracting the smallest bid and dividing by the number of remaining participants. The most common mistake here was to assume that the maximum allocation calculated in part (b) still applied (yielding 20, 26, 39, 39).
- (d) A level of abstraction was required to tackle this question – appreciating that the two tokens unsold must involve three players getting the maximum allocation. Trial and improvement could then lead to a correct lowest bid to achieve the required outcome. Few candidates managed this correctly.
- (e) This question was orientated towards the price-setting aspect of the game, when combined with the increase in tokens each round. Success at this question was independent of success at parts (b), (c) and (d). Trial and improvement could be used to find the number of tokens that produced the limit value of \$10, and hence deduce the correct strategy for Eliza. Follow-through marks were available for those who answered (a) incorrectly.
- (f) The change of price by \$5.50 was selected to be challenging: the increase in tokens must be less than 10 (since the game has not ended) and therefore the prices must straddle the \$50 point. Trial and improvement (or logical deduction!) yields the options given in the mark scheme. The logical inter-dependencies in this question clearly defeated most candidates.
- (g) This tactical question proved to be too much for most candidates. It depended upon a grasp of the price-setting mechanism investigated in (e) and (f), combined with an appreciation of how to keep the game going at the brink. The most common attempt at this question was simply to state how many tokens Jill could buy in one round.

Question 4

This question involved the interplay of data from a sporting league table, scores from yesterday and matches to be played in the final forthcoming week – all to be analysed in the light of a scoring mechanism, made up of two constituents, which involved some subtle logical interdependencies. All of the questions were independent of each other – each acting as a separate test of the implications and interdependencies that existed in the data. There was an impression that many candidates would have benefitted from studying the information carefully, rather than simply reading it quickly, before attempting the questions.

- (a) The first question directed candidates to one of the limited cases where the identity of the team scoring the first goal could be reliably identified; but many candidates did not appreciate this, awarding the Jays 3 points rather than 4. This question was answered correctly by disappointingly few candidates.
- (b) The three different time lines in this question were the source of a number of potential misconceptions – and the limited success that candidates had with this question implied that many failed to draw the appropriate inferences from the fact that the question referred to the status quo before yesterday. The most common errors were to classify all of the Jays' non-wins as losses, or to ignore the Nutcrackers' loss yesterday.
- (c) A sure answer to this question depended on calculating how many of the Ravens' 36 points were achieved by other means (losses, draws and first goals): the only subtlety in this was to appreciate that the size of their loss was known, since it occurred yesterday. The majority of candidates were able to deduct the points from the draws and first goals (gaining a mark for this) and made a guess with respect to the loss.
- (d) Success in this question depended on a realisation that 0–0 draws were the only matches in which there was no first goal. Few candidates saw this.
- (e) The restrictions that circumscribed the information in this question were entirely independent of the timelines and the other data in the table. Most candidates were able to identify scores that summed to 7 with the appropriate allocation of points.
- (f) A detailed explanation supported by evidence from the table was required to gain full marks here, and very few candidates managed this. The analysis needed to show that the best outcomes for the Jackdaws, and the worst for the other contending teams, would still not lead to their overall victory. Many candidates failed to specify that the best outcome for the Jackdaws would leave them with 37 points. And those who did tended to finish their analysis with the fact that they could then be equal with the Jays and Choughs – very few realising that an appeal to First Goal points decided the issue.
- (g) The three restrictions (Ravens win overall, fewer matches won, independent of first goals) place different restrictions on the scores: most candidates were able to see that the Rooks and Crows had to win to ensure that the Ravens had fewer wins than any other team. That both of the other matches had to be drawn was seen by some candidates – with only a small number appreciating that a 0–0 draw was necessary to avoid the Jays or Choughs gaining an extra point.
- (h) This question was independent of all previous parts – and involved a reconsideration of how many matches are played between n teams. 8 teams play $(7+6+5+4+3+2+1)$ games, twice. 10 teams play $(9+8+7+6+5+4+3+2+1)$ games, twice. A number of candidates took the question to mean 'how many more games would be played by each team', but were awarded 1 mark for this.

THINKING SKILLS

Paper 9694/32
Problem Analysis
and Solution

Key Messages

Candidates should show key elements of their working for any question which offers more than 1 mark. Where an explanation is asked for, at least one sentence is required. Many could get higher marks by detecting and correcting errors that led to implausible answers.

General Comments

Most candidates obtained some marks on all four questions. There was more time this year than previously, and so more complete answers. Candidates did not waste time on poetry or irrelevant graphics, and were generally better prepared than previously.

It is normally advisable to use the units as given to avoid errors in double conversion, e.g. easy arithmetic for cars travelling in km/h will be harder in m/s, even if that is the SI unit.

Fractions should be treated as such and not (inconsistently) rounded or turned into percentages, unless this is specifically asked for.

Answers that are completely out of range should ring alarm bells: is it plausible that a bus ticket will be \$168.42 when the others are \$4.30 and \$3.20?

Questions are formed of parts which are not randomly selected: candidates should consider whether a part is exploring more about something in the previous part or is looking at an alternative process or aspect.

To gain partial credit for incomplete work and full credit for correct work, candidates should only cross out responses when they have replaced them.

Comments on Specific Questions

Question 1

This question required candidates to select products from a table according to different criteria, to propose a better way for the data to be displayed, and to construct an example of another product that could not be competitive under any reasonable combined criteria with any of the given ones, but also always better than another in some aspect. Candidates were asked both to find what people with different priorities would select and to construct an example that would not be chosen, whatever a person thought was important. They were not required to devise their own scoring system, nor to make judgements on value for money.

- (a) The simple extraction from the table of the best choice for a single criterion was correctly performed by almost all candidates.
- (b) This part required candidates to show that there were no options available for optimisation of two different requirements, and most candidates were able to identify why there was no way to satisfy both.

- (c) This required finding the cases where two constraints were satisfied, and then selecting the optimal value using secondary considerations. Most candidates found the best case, but many offered their own algorithm (such as using average ratings or some value for money formula), rather than what they were told the customer wanted.
- (d) This part called for a reversal of the display of one of the components because it was the only one where it was better to have a smaller number. Any transformation that provided a reversal was acceptable.
- (e) Although drawing two pentagons on one set of axes was asked for (so that they could be compared), half of the candidates used two sets.
- (f) This called for creation of an example and then identification of one of the existing buckets that was as good or better in all aspects. Most candidates offered an example, although some forgot that the best price was lowest, but few identified a specific existing bucket that was as good as or better in all respects. Higher marks were scored by those who understood that the previous parts of the question offered helpful hints or reminders.

Question 2

This required finding the correct combinations for extracting personal information from simple statistics on a very small data set, and then using the techniques developed on a similar second data set. It called for logical deduction, with minimal extra arithmetic, but did require candidates to be familiar with the definition of a median, and the convention for calculating medians on even-sized sets.

- (a) Candidates were asked to offer an alternative hypothesis, but very few appreciated that meant assuming that the hypothesis given was false, and that the medians might be the result of the heights of (one or) two other suspects. The specific case of a match was not a suggested general method for later use, as it was presented as a hypothesis that might be wrong.
- (b) This part called for the selection of a suitable small set to use, and then the deduction of a height from it. This can be done since the median and mean are the same when there are fewer than three items.
- (c) Using the result of (b), a similar approach allows the remaining heights to be determined. There are three ways to do this, all equally creditworthy.
- (d) This part sought an explanation (in words) of why the information given was not sufficient to deduce a particular value.
- (e) This involved using the process developed earlier on a new set of data. Some candidates who had correctly pointed out why one of the values could not be determined in (d) wasted time by attempting to include this value.

Question 3

This question involved deducing times, distances and prices from information about a real but simplified bus route, offering a hypothesis for features of the rules, and the comparison of alternative pricing models. It required particular care about which end of a zone was being considered.

- (a) This part asked for the number of types of buses that could be used for a particular journey. Most candidates correctly determined which types, but many omitted to count them, which was what the question asked for. A few wasted time after reaching the answer, e.g. by attempting to multiply the number of types by the maximum number of passengers on a bus. It always helps to check that the units of an answer match those of the question.
- (b) The calculation of the maximum possible number of fares required appreciation both of the constant in-zone fare and that fares between zones could all be different. Very few gained the two marks for 1 (for travel inside a zone) + choosing any 2 from 9 ($8 + 7 + 6 + 5 + 4 + 3 + 2 + 1$).

- (c) This asked for a plausible reason for what is, at first sight, a strange rule that could result in different prices in different directions on the same bus. Any hypothesis that didn't contradict the assumptions and wouldn't make something worse was acceptable, whether general or described by an example of a specific case.
- (d) This required extraction of the relevant information from the table, combined with the restriction to travel during daylight hours, using the 24 hour clock. It required care about which end of a zone was being considered, and that there are 60 rather than 100 minutes in an hour.
- (e) This required a distance to be determined from the time of travel. Some candidates converted speeds to metres/second, making the calculations much harder and often introducing errors.

Since zones vary in length, the first part looked for addition of the first four lengths.

The second part gave an answer and asked how it was derived. This could be done either by considering the times or previously calculated distances. Candidates should be aware that there is no credit for just the answer when that is given in the question.

- (f) This required determining the parameters in a simple formula and then applying it to a specific case. There were partial marks available when the working was shown, even where common arithmetic errors were made. Although the type of dollar is not specified, a few candidates should have checked that their answers were in a plausible range compared with the examples given.
- (g) This explored what would happen for the same case under an alternative simpler proposal. Candidates should note that it would be unlikely for exactly the same question to be asked twice, so 'same as f' is unlikely to be correct and to gain 2 marks.

Question 4

This question required candidates to consider general features of a novel game in order to find the best strategy for players in specific cases, to deduce information about where certain tiles must be, and to calculate the theoretical highest possible score. Many candidates took insufficient care about the distinctions between numbers on tiles, totals and scores. It was important to note that the distance from 150 was important; it was not a threshold.

- (a) This required calculation of both players' totals in a particular scenario, and deduction as to which one was the winner. It was not sufficient to write down the winner. The second part involved understanding the distinction between the totals and the score.
- (b) This part explored the best available options in three specific cases. The cases are superficially similar, but many candidates only found two correctly, usually by mistaking 150 as a threshold. A few offered value judgements such as 'too far away' instead of numerical answers, despite having done the relevant calculations.
- (c) This part sought the best strategy by considering the opponent's best options in the same three cases, and explaining why this was sure to win.
- (d) This part required determining various sets of tiles, and then using arithmetic to determine the only possible unseen pair. Candidates needed to use the fact that the sum of all the tiles is constant.
- (e) The last part required demonstration of how the (given) maximum score could be achieved. Not only did the highest number tiles need to be used; the bonus for getting exactly 150 was also needed.

THINKING SKILLS

Paper 9694/33
Problem Analysis
and Solution

Key Messages

Candidates should show key elements of their working for any question which offers more than 1 mark. Where an explanation is asked for, at least one sentence is required. Many could get higher marks by detecting and correcting errors that led to implausible answers.

General Comments

Most candidates obtained some marks on all four questions. There was more time this year than previously, and so more complete answers. Candidates did not waste time on poetry or irrelevant graphics, and were generally better prepared than previously.

It is normally advisable to use the units as given to avoid errors in double conversion, e.g. easy arithmetic for cars travelling in km/h will be harder in m/s, even if that is the SI unit.

Fractions should be treated as such and not (inconsistently) rounded or turned into percentages, unless this is specifically asked for.

Answers that are completely out of range should ring alarm bells: is it plausible that a bus ticket will be \$168.42 when the others are \$4.30 and \$3.20?

Questions are formed of parts which are not randomly selected: candidates should consider whether a part is exploring more about something in the previous part or is looking at an alternative process or aspect.

To gain partial credit for incomplete work and full credit for correct work, candidates should only cross out responses when they have replaced them.

Comments on Specific Questions

Question 1

This question required candidates to select products from a table according to different criteria, to propose a better way for the data to be displayed, and to construct an example of another product that could not be competitive under any reasonable combined criteria with any of the given ones, but also always better than another in some aspect. Candidates were asked both to find what people with different priorities would select and to construct an example that would not be chosen, whatever a person thought was important. They were not required to devise their own scoring system, nor to make judgements on value for money.

- (a) The simple extraction from the table of the best choice for a single criterion was correctly performed by almost all candidates.
- (b) This part required candidates to show that there were no options available for optimisation of two different requirements, and most candidates were able to identify why there was no way to satisfy both.

- (c) This required finding the cases where two constraints were satisfied, and then selecting the optimal value using secondary considerations. Most candidates found the best case, but many offered their own algorithm (such as using average ratings or some value for money formula), rather than what they were told the customer wanted.
- (d) This part called for a reversal of the display of one of the components because it was the only one where it was better to have a smaller number. Any transformation that provided a reversal was acceptable.
- (e) Although drawing two pentagons on one set of axes was asked for (so that they could be compared), half of the candidates used two sets.
- (f) This called for creation of an example and then identification of one of the existing buckets that was as good or better in all aspects. Most candidates offered an example, although some forgot that the best price was lowest, but few identified a specific existing bucket that was as good as or better in all respects. Higher marks were scored by those who understood that the previous parts of the question offered helpful hints or reminders.

Question 2

This required finding the correct combinations for extracting personal information from simple statistics on a very small data set, and then using the techniques developed on a similar second data set. It called for logical deduction, with minimal extra arithmetic, but did require candidates to be familiar with the definition of a median, and the convention for calculating medians on even-sized sets.

- (a) Candidates were asked to offer an alternative hypothesis, but very few appreciated that meant assuming that the hypothesis given was false, and that the medians might be the result of the heights of (one or) two other suspects. The specific case of a match was not a suggested general method for later use, as it was presented as a hypothesis that might be wrong.
- (b) This part called for the selection of a suitable small set to use, and then the deduction of a height from it. This can be done since the median and mean are the same when there are fewer than three items.
- (c) Using the result of (b), a similar approach allows the remaining heights to be determined. There are three ways to do this, all equally creditworthy.
- (d) This part sought an explanation (in words) of why the information given was not sufficient to deduce a particular value.
- (e) This involved using the process developed earlier on a new set of data. Some candidates who had correctly pointed out why one of the values could not be determined in (d) wasted time by attempting to include this value.

Question 3

This question involved deducing times, distances and prices from information about a real but simplified bus route, offering a hypothesis for features of the rules, and the comparison of alternative pricing models. It required particular care about which end of a zone was being considered.

- (a) This part asked for the number of types of buses that could be used for a particular journey. Most candidates correctly determined which types, but many omitted to count them, which was what the question asked for. A few wasted time after reaching the answer, e.g. by attempting to multiply the number of types by the maximum number of passengers on a bus. It always helps to check that the units of an answer match those of the question.
- (b) The calculation of the maximum possible number of fares required appreciation both of the constant in-zone fare and that fares between zones could all be different. Very few gained the two marks for 1 (for travel inside a zone) + choosing any 2 from 9 ($8 + 7 + 6 + 5 + 4 + 3 + 2 + 1$).

- (c) This asked for a plausible reason for what is, at first sight, a strange rule that could result in different prices in different directions on the same bus. Any hypothesis that didn't contradict the assumptions and wouldn't make something worse was acceptable, whether general or described by an example of a specific case.
- (d) This required extraction of the relevant information from the table, combined with the restriction to travel during daylight hours, using the 24 hour clock. It required care about which end of a zone was being considered, and that there are 60 rather than 100 minutes in an hour.
- (e) This required a distance to be determined from the time of travel. Some candidates converted speeds to metres/second, making the calculations much harder and often introducing errors.

Since zones vary in length, the first part looked for addition of the first four lengths.

The second part gave an answer and asked how it was derived. This could be done either by considering the times or previously calculated distances. Candidates should be aware that there is no credit for just the answer when that is given in the question.

- (f) This required determining the parameters in a simple formula and then applying it to a specific case. There were partial marks available when the working was shown, even where common arithmetic errors were made. Although the type of dollar is not specified, a few candidates should have checked that their answers were in a plausible range compared with the examples given.
- (g) This explored what would happen for the same case under an alternative simpler proposal. Candidates should note that it would be unlikely for exactly the same question to be asked twice, so 'same as f' is unlikely to be correct and to gain 2 marks.

Question 4

This question required candidates to consider general features of a novel game in order to find the best strategy for players in specific cases, to deduce information about where certain tiles must be, and to calculate the theoretical highest possible score. Many candidates took insufficient care about the distinctions between numbers on tiles, totals and scores. It was important to note that the distance from 150 was important; it was not a threshold.

- (a) This required calculation of both players' totals in a particular scenario, and deduction as to which one was the winner. It was not sufficient to write down the winner. The second part involved understanding the distinction between the totals and the score.
- (b) This part explored the best available options in three specific cases. The cases are superficially similar, but many candidates only found two correctly, usually by mistaking 150 as a threshold. A few offered value judgements such as 'too far away' instead of numerical answers, despite having done the relevant calculations.
- (c) This part sought the best strategy by considering the opponent's best options in the same three cases, and explaining why this was sure to win.
- (d) This part required determining various sets of tiles, and then using arithmetic to determine the only possible unseen pair. Candidates needed to use the fact that the sum of all the tiles is constant.
- (e) The last part required demonstration of how the (given) maximum score could be achieved. Not only did the highest number tiles need to be used; the bonus for getting exactly 150 was also needed.

THINKING SKILLS

Paper 9694/41
Applied Reasoning

Key Messages

- The first question in this paper tested the candidates' ability to evaluate claims based on statistical data.
- In **Question 2** candidates had the opportunity to display their ability to analyse the structure of a reasoned argument.
- In **Question 3** candidates only gained marks if they identified weaknesses in the reasoning within the document.
- **Question 4** allowed candidates to use a full range of critical reasoning skills in order to construct a reasoned argument using information from the documents.

General Comments

There was little evidence of candidates running out of time on this paper. The number of candidates writing answers whose length does not reflect the mark allocation is decreasing but still significant.

The standard of candidates varied greatly. There was evidence that many candidates had been taught some of the language of reasoning and some were familiar with the format of the paper. Indeed, some candidates appeared particularly well-prepared in this regard and answered **Question 4** first, attempting to ensure that the most creditworthy question was not rushed. While there are merits to this strategy, it is often useful to tackle **Questions 2** and **3** before **Question 4** in order to develop a deeper understanding of the topic, and the structure and shortcomings of Document 1.

Comments on Specific Questions

Question 1

- (a) Candidates were asked to criticise a report about crime figures in the UK in comparison with other countries. Many candidates gained one mark and some gained two. The most common correct responses were that population was not taken into account and that France and Austria might have been selected to exaggerate differences. Fewer candidates commented on the degree of violence or the possible different definitions of 'crime' and 'violent' in different countries. A large number of candidates questioned the source of the data or criticised a perceived lack of numerical detail, for which no marks were available.
- (b) Candidates were asked to comment on how well a politician's claim was supported by the crime report. Many candidates gained one mark for questioning the assumed link between violent crime and social problems, but few phrased this in terms of an assumption to gain the second mark. Likewise, some candidates gained one mark for a discussion of the unknown effect of any previous government intervention, but answers were not sufficiently well expressed to gain two marks. Very few candidates noticed that the figures only covered one year and so a trend could not be inferred; those few that did gained both marks.

Question 2

Candidates who had been prepared for the examination usually gained some marks for this question, often for identifying one or more of the intermediate conclusions or counter-assertions. Very few recognised the main conclusion. As always, some candidates did not understand what was required of the task and attempted to paraphrase, summarise or criticise the argument. Successful candidates identified parts of the text, copied them out and labelled them as MC, IC, etc.

Question 3

Successful evaluation of the argument was possible for those who knew what was expected. Those candidates who had learned some critical thinking tended to score between two and five marks. Where marks were awarded, it was usually for identifying the *ad hominem* attack on the tree-huggers or for questioning irrelevant comparisons with, or assumptions about, air freight. Other marking points were awarded, but infrequently. As ever, many candidates simply stated a series of counter-assertions to the claims made in the document and received no credit.

Question 4

The majority of candidates clearly understood the topic but found it difficult to introduce ideas of their own. The fact that most documents supported the given conclusion helped the weaker candidates construct focused arguments. As ever, many made no critical comments about the documents they were using, and in this series few candidates seemed to be using one document to contrast with or support another. Thus marks in the 'use of documents' skill area were often limited. It was good that few candidates answered the question by discussing each document in turn, rather than structuring their answer as an argument, which meant that it was often easier to award marks in the 'structure' skill area. It was also good to see the majority of candidates offer some sort of conclusion, with fewer avoiding the conclusion they are given and instead arguing for a 'soft' compromise position. The fact that candidates struggled to introduce their own ideas meant that marks were sometimes limited, not only in the 'quality of argument' skill, but also in the skill area covering the treatment of counter-positions.

THINKING SKILLS

Paper 9694/42
Applied Reasoning

Key Messages

- The first question in this paper tested the candidates' ability to evaluate a claim based on statistical data.
- In **Question 2** candidates had the opportunity to display their ability to analyse the structure of a detailed argument.
- In **Question 3** candidates only gained marks if they identified weaknesses in the reasoning within the document.
- **Question 4** allowed candidates to use a full range of critical reasoning skills in order to construct a reasoned argument using information from the documents.

General Comments

There was little evidence that candidates were running out of time on this paper. The number of candidates writing answers whose length does not reflect the mark allocation is decreasing but still significant.

The standard of candidates was somewhat higher than in recent series, particularly on 9694/42, and, for many candidates, there was evidence of familiarity with the language of reasoning and the format of the paper.

Comments on Specific Questions

Question 1

The majority of candidates attempted to criticise the passage, with few straying beyond the statistics themselves. Most candidates scored two or three marks, but the full range of available marks was seen. The most common correct responses alluded to the unrepresentativeness of the sample, the potential differences in the number of male and female respondents, and the paucity of choices presented. Fewer candidates questioned the vertical scale on the chart, the size of the actual difference or the meanings of 'enjoy', 'watch' and 'watch more'.

Question 2

The majority of candidates seemed aware of what was required for this task. Successful candidates identified parts of the text, copied them out and labelled them as MC, IC, or CA. Most candidates struggled to identify the main conclusion, but often picked up marks for identifying one or more intermediate conclusions or counter-assertions. Fewer candidates than usual summarised the argument or provided paraphrases of the key structural elements.

Question 3

The argument itself was challenging to evaluate, but marks were available for those who had been trained in critical evaluation. The majority of candidates, however, scored very few marks. Many merely listed counter-assertions to points given in the document. Fewer than usual cited a lack of supporting evidence, statistics or source identification, suggesting that they had received some training in critical evaluation. The most common creditworthy responses cited circularity or question-begging, the straw man reference to 'little green men', or the frequent use of loaded or rhetorical language. Marks were awarded for all the other marking points, but infrequently.

Question 4

The majority of candidates appeared to find the subject matter accessible and this allowed them to structure their own arguments on the topic and bring in ideas of their own. A small number of candidates produced excellent responses which scored highly in all four skill areas. Fewer than usual candidates merely listed the points and the documents supporting each side and did not make a reasoned case. This meant that it was easier to award marks for structure. Likewise, it was, in many cases, possible to award marks for treatment of counter-positions as many candidates were able to recognise and discuss realistic counter-positions. While the vast majority of candidates used the documents to support their argument, many made no critical comments about the documents they were using, limiting their mark in the 'use of documents' skill area. Most candidates offered a reasonable conclusion.

THINKING SKILLS

Paper 9694/43
Applied Reasoning

Key Messages

- The first question in this paper tested the candidates' ability to evaluate a claim based on statistical data.
- In **Question 2** candidates had the opportunity to display their ability to analyse the structure of a detailed argument.
- In **Question 3** candidates only gained marks if they identified weaknesses in the reasoning within the document.
- **Question 4** allowed candidates to use a full range of critical reasoning skills in order to construct a reasoned argument using information from the documents.

General Comments

There was little evidence that candidates were running out of time on this paper. The number of candidates writing answers whose length does not reflect the mark allocation is decreasing but still significant.

The standard of candidates was somewhat higher than in recent series, particularly on 9694/42, and, for many candidates, there was evidence of familiarity with the language of reasoning and the format of the paper.

Comments on Specific Questions

Question 1

The majority of candidates attempted to criticise the passage, with few straying beyond the statistics themselves. Most candidates scored two or three marks, but the full range of available marks was seen. The most common correct responses alluded to the unrepresentativeness of the sample, the potential differences in the number of male and female respondents, and the paucity of choices presented. Fewer candidates questioned the vertical scale on the chart, the size of the actual difference or the meanings of 'enjoy', 'watch' and 'watch more'.

Question 2

The majority of candidates seemed aware of what was required for this task. Successful candidates identified parts of the text, copied them out and labelled them as MC, IC, or CA. Most candidates struggled to identify the main conclusion, but often picked up marks for identifying one or more intermediate conclusions or counter-assertions. Fewer candidates than usual summarised the argument or provided paraphrases of the key structural elements.

Question 3

The argument itself was challenging to evaluate, but marks were available for those who had been trained in critical evaluation. The majority of candidates, however, scored very few marks. Many merely listed counter-assertions to points given in the document. Fewer than usual cited a lack of supporting evidence, statistics or source identification, suggesting that they had received some training in critical evaluation. The most common creditworthy responses cited circularity or question-begging, the straw man reference to 'little green men', or the frequent use of loaded or rhetorical language. Marks were awarded for all the other marking points, but infrequently.

Question 4

The majority of candidates appeared to find the subject matter accessible and this allowed them to structure their own arguments on the topic and bring in ideas of their own. A small number of candidates produced excellent responses which scored highly in all four skill areas. Fewer than usual candidates merely listed the points and the documents supporting each side and did not make a reasoned case. This meant that it was easier to award marks for structure. Likewise, it was, in many cases, possible to award marks for treatment of counter-positions as many candidates were able to recognise and discuss realistic counter-positions. While the vast majority of candidates used the documents to support their argument, many made no critical comments about the documents they were using, limiting their mark in the 'use of documents' skill area. Most candidates offered a reasonable conclusion.