



Office of  
the Schools  
Adjudicator

## Determination

<b>Case references:</b>	<b>ADA4281-4293</b>
<b>Objector:</b>	<b>Two members of the public</b>
<b>Admission authorities:</b>	<b>The admission authorities for the 13 grammar schools in Buckinghamshire</b>
<b>Date of decision:</b>	<b>16 October 2024</b>

## Determination

**In accordance with section 88H(4) of the School Standards and Framework Act 1998, I do not uphold the objections to the admission arrangements for September 2025 determined by the admission authorities for the 13 grammar schools in Buckinghamshire.**

**I have also considered the various arrangements in accordance with section 88I(5) and find there are other matters which do not conform with the requirements relating to some of the admission arrangements in the ways set out in this determination.**

**By virtue of section 88K(2) the adjudicator's decision is binding on the relevant admission authorities. The School Admissions Code requires those admission authorities to revise their admission arrangements within the time scales set out in this determination.**

## The referral

1. Under section 88H(2) of the School Standards and Framework Act 1998 (the Act), an objection has been referred to the adjudicator by two members of the public (the objectors) about the admission arrangements (the arrangements) for the 13 grammar schools in Buckinghamshire, known collectively as The Buckinghamshire Grammar Schools (TBGS). The schools were all designated as grammar schools under Section 104(5) of the School Standards and Framework Act 1998, have all converted to academy status, and all cater for children aged 11-18. They have set up TBGS as a company in order to operate a coordinated selection system using a common secondary transfer test.

2. The objection is complex and detailed, but its essence is that the arrangements as currently determined are not compliant with the School Admissions Code (the Code). The objectors are of the view that the arrangements do not contain sufficient information necessary to comply with the requirement of clarity set out in paragraph 14 of the Code, which states that:

“...admission authorities **must** ensure that the practices and the criteria used to decide the allocation of school places are fair, clear, and objective. Parents should be able to look at a set of arrangements and understand easily how places for that school will be allocated.”

They also assert that the arrangements operate unfairly.

3. The local authority for the area in which the schools are located is Buckinghamshire County Council (the LA). The LA is a party to the objection, as are TBGS, the objectors, the LA, and the various trusts who are the admission authorities for the schools that make up TBGS.

## Jurisdiction

4. The terms of the academy agreements between the various trusts and the Secretary of State for Education require that the admissions policies and arrangements for the academy schools are in accordance with admissions law, in some cases as it applies to foundation and voluntary aided schools, and in the rest as it applies to maintained schools. These arrangements were determined by the trusts, which are the individual admission authorities for the schools, on that basis. The objectors submitted the objection to these determined arrangements on 29 March 2024. I am satisfied that the objection has been properly referred to me in accordance with section 88H of the Act and that it is within my jurisdiction. For 10 of the 13 sets of arrangements, I have also used my power under section 88I of the Act to consider the arrangements as a whole.

## Procedure

5. In considering this matter I have had regard to all relevant legislation and the Code.

6. The documents I have considered in reaching my decision include:

- a. copies of the minutes of the meetings of the various trust boards at which their arrangements were determined;
- b. copies of each of the determined arrangements;
- c. the objectors' form of objection dated March 2024 and supporting documents;
- d. the response from TBGS to the objection, supporting documents and subsequent correspondence (the chair of TBGS was nominated by each of the schools to act as a central point of contact and authorised by each to speak and act on their behalf);

- e. a paper entitled “Reliability, Validity and all that Jazz”, written in 2001 by Dylan William, then Assistant Principal of Kings College, London;
- f. rulings from a number of tribunals in relation to FOI requests made by the objectors;
- g. various other sources of information about reliability, which are listed in the relevant section of this determination;
- h. further correspondence and supporting documents from TBGS and the objectors, in response to my additional enquiries; and
- i. information available on the websites of the local authority, the schools, TBGS and the Department of Education (DfE).

## The Objection

7. The objection was submitted on the standard objection form, along with a detailed supporting case document in excess of 12,000 words. In essence the objection as originally submitted can be summarised as follows:

- a. that the published arrangements should include “an unequivocal commitment to disclose sufficient information so that the selection process can be subjected to legitimate public scrutiny. “
- b. that the arrangements should provide for certain data from the tests to be routinely published, such as “an independently verified test/retest reliability index, confidence intervals and the probability of misclassifications”. Alternatively, the arrangements should provide for suitable raw data to be made available for others to analyse in a similar fashion.
- c. that because there is a “zone of uncertainty” for each child’s score, there should be some mechanism for expressing that to parents, for use in the Selection Review process and/or independent appeal process. The objectors argue that the arrangements should ensure that the remit of the Selection Review is “explicitly articulated to include a properly understood approach to appeals from those parents whose child’s performance falls within the ‘zone of uncertainty.’”
- d. that the current process of testing as outlined in the arrangements is intrinsically unfair and “could be improved”, for example by administering tests twice to each candidate and the better performance being used to decide eligibility.

8. The objection documents and subsequent correspondence show that there has been a long history of dealings between the objectors and TBGS, which include Freedom of Information (FOI) requests, referrals to the Information Commissioner’s Office, and appeals to the First Tier Tribunal. After reading all the extensive material provided, both in the

original objection documents and in subsequent correspondence, I asked the objectors to clarify their position and they have done so as follows:

“...[we] are of the view that the arrangements cannot be fully compliant with the Code’s requirement for fairness, unless they make the methodology used to translate raw test scores into a pass/fail decision fully transparent. The methodology should be set out in general terms on the face of the arrangements and parents should be given the outcome of a recognised ‘test/retest reliability’ measure calculated for that iteration of the test, alongside the results for each child. Additionally, the arrangements should confirm that the admissions authority would be willing to respond positively to any legitimate request for a complete anonymised dataset comprising all raw test scores and a measure of age for each child, together with full details of the algorithms used to translate these scores into the standardised and scaled results as well as the reliability calculation communicated to parents, in order to allow third parties to analyse and verify any claims made by TBGS.”

In order to make my determination easier to follow, I have therefore re-cast the wording in the objection and the clarification statement into four parts:

Part 1 – reliability and intrinsic unfairness. In this section, I will address the matters of intrinsic unfairness and its relationship to the reliability of the test. I will consider some of the more detailed issues about reliability in part 4.

Part 2 – the proposed revised reporting of results. In this section I will address the impact of reliability on an individual test score and the objectors view that each child’s test score should be reported with a confidence interval and/or a misclassification probability.

Part 3 - the proposed improvements to the testing process. In this section I will address the suggestion about testing candidates more than once.

Part 4 – the remaining proposed additions. In this section, I will address other matters that relate to what the objectors say needs to be added to the arrangements, such as full details of the methodology used to turn raw answer data into the binary 11+ status, the age-standardisation and scaling algorithms, and a commitment to releasing a complete anonymised data set on request to any third party. In this section I will also address the matter of reliability in more detail.

9. Paragraph 14 of the Code requires that arrangements are fair, clear and objective. There is no suggestion in the objection that the objectors have concerns about the objectivity of the arrangements, and therefore my considerations will focus on whether the arrangements are fair and clear.

## Other Matters

10. I have also considered that in 10 of the 13 sets of arrangements, it appeared to me that the description of the “Selection Review” process might not be sufficiently clear to comply with the requirements of paragraph 14 of the Code.

## Background

### TBGS

11. The following table sets out the schools that comprise TBGS together with their admission authorities:

**Table 1: TBGS schools**

School	Gender of Entry	Admission Authority
Aylesbury Grammar School	Boys	Governing body (single-academy trust)
Aylesbury High School	Girls	Governing body (single-academy trust)
Beaconsfield High School	Girls	Governing body (single-academy trust)
Burnham Grammar School	Mixed	Beeches Learning and Development Trust
Chesham Grammar School	Mixed	Red Kite Schools Trust
Dr Challoner's Grammar School	Boys	Governing body (single-academy trust)
Dr Challoner's High School	Girls	Governing body (single-academy trust)
John Hampden Grammar School	Boys	Governing body (single-academy trust)
Royal Latin School	Mixed	Governing body (single-academy trust)
Sir Henry Floyd Grammar School	Mixed	Insignis Academy Trust
Sir William Borlase's Grammar School	Mixed	Marlow Education Trust
The Royal Grammar School, High Wycombe	Boys	Governing body (single-academy trust)
Wycombe High School	Girls	Wycombe High School Academies Trust

12. The TBGS website explains the relationship between TBGS and the above academies as follows:

“The Buckinghamshire Grammar Schools (TBGS) is a company set up by the thirteen grammar schools in Buckinghamshire in order to manage and administer The Buckinghamshire Secondary Transfer Test.

All the Buckinghamshire grammar schools are academies and therefore their own admissions authorities, but following their academy conversion the thirteen schools were very keen to continue their long tradition of working together. They wanted to maintain a coordinated selection system using a common secondary transfer test to select pupils for entry into all Buckinghamshire grammar schools at 11+ each year, to avoid an admissions and testing ‘free for all’ which would be to the detriment of children and families. In order to do this, TBGS was established.”

13. Leaving aside the statutory independent admissions appeal process, there are two ways in which a child can be considered to be qualified for admission to a TBGS school. Either they must achieve the required qualifying score of 121 in the Buckinghamshire Secondary Transfer Test (STT), or, if they do not achieve that score, they can be deemed qualified by a process called Selection Review (SR). Once deemed qualified, any such child is then eligible for consideration for entry to any of the TBGS schools, although each school then applies its own oversubscription criteria and a place at a TBGS school is not guaranteed.

14. It is important to note that once a child is deemed to be qualified, either by scoring at least 121 or by the SR process, the actual score achieved plays no further part in the process of application to TBGS schools nor in their oversubscription criteria.

#### Qualification via testing

15. The testing process is relatively straightforward. Each year around 10,000 children sit the test in the autumn term of year 6 and the majority do so at their own primary school. The test is commissioned by TBGS acting on behalf of the 13 Buckinghamshire grammar schools and comprises two papers that are each approximately 60 minutes in length and both papers are taken on the same day with a short gap in between. One paper assesses verbal skills, whilst the other assesses mathematical and non-verbal skills.

16. The test is produced and marked by GL Assessment, who also age-standardise the results and analyse them on behalf of schools. The website of the company describes itself as “the leading provider of formative assessments to schools across the UK and Ireland”.

17. A child’s scores in each of the three skills areas are age-standardised and a weighted sum of the scores is created. The weightings are 50% verbal skills, 25% mathematical skills, 25% non-verbal skills. A scaling process is also incorporated into the

standardisation such that a final overall score of 121 is deemed to be a “qualifying” score. TBGS tell me that:

“The model used is designed to support the number of places available in Buckinghamshire grammar schools.”

“A summary of the draft results is shared with the TBGS data group to inform the final decision regarding the number of pupils who qualify for a grammar school place in the test and GL Assessment then uses this decision to generate the results.”

### Qualification via Selection Review

18. The SR process is outlined in detail on the website of the LA. I reproduce much of it here at length because the details are relevant to some of the matters raised in the objection.

“On each Selection Review Panel (SRP) there are two grammar school headteachers and one primary school headteacher.

All headteachers who sit on the panels receive training. They understand the nature of the decisions they are taking. They will not consider cases of children known to them or from schools in which they have an interest. The SRPs are not told which secondary schools you have applied for.

Sometimes, the SRP will be joined by an Educational Psychologist (EP), or similar education professional with an understanding of the technical content of specialist reports. This will happen if a child’s case includes an Educational Psychologist report, or another specialist report (e.g. from a medical consultant) where it is felt that a specialist’s input would be beneficial. If your child was considered for reasonable adjustments to the Secondary Transfer Test, their Selection Review will also be heard by a panel supported in this way.

[...]

A successful Selection Review will usually include strong academic evidence that shows your child is working to the level expected of a potential grammar school pupil, as well as evidence of exceptional reasons showing why your child did not perform as well as expected in the test.

The purpose of the evidence is to demonstrate to the Selection Review Panel (SRP) that your child’s results were not as expected: in other words, your child would have qualified, if not for the exceptional circumstances.

Academic evidence could include (but is not limited to):

- The Headteacher’s Selection Review Summary Sheet
- School reports

- Supporting letters from your child's class teacher

Please do not include examples of school work, as the SRP will not be able to consider this. Evidence of participation in extra-curricular activities, whilst commendable, is also not information the SRP will take into account.

Evidence of exceptional reasons for not qualifying could include (but is not limited to):

- A medical letter (from a GP or hospital) in relation to illness of your child or a close family member
- A letter from an employer to explain a lengthy parental absence (particularly for forces families where a family member has been serving in a combat zone)
- Evidence that shows significant issues – such as the death of a family member – near to the time of the test
- A very recent house or school move

The Selection Review Panel (SRP) will be provided with your child's final score, as well as their scores in the verbal, mathematical and non-verbal sections of the test.

If your child attends a state funded Buckinghamshire primary school, or a Partner school, we will also provide the SRP with a recommendation from your child's headteacher.

Headteachers of state funded Buckinghamshire primary or Partner schools are asked – before the test results are known – whether they would recommend your child for a grammar school. They also provide information about your child's attitude to learning. If your child attends any other school, your child's headteacher will have the opportunity to add their recommendation as part of the Headteacher's Selection Review Summary Sheet.

The headteacher can give a child's grammar school suitability recommendation as:

- (1) Exceptionally able so very highly recommended, or;
- (2) Very able so recommended without any reservation, or;
- (3) Recommended with reservation, or;
- (4) Not recommended for grammar school.

Your child's attitude to learning will be described as:

- (1) Enjoys a challenge and is a highly motivated independent learner, or;
- (2) Consistently hard working and reliable, or;



(3) Output varies, or;

(4) Lacks self-organisation, requires support.

This information is valued by the SRP, because it gives them an idea of the context in which your child is working in school. It can also help to show if a child's result was unexpected when compared to their recommendation.

[...]

If your child's Selection Review is successful, this means that your child is now treated the same as a child who scored 121 or more in the Secondary Transfer Test: they are considered qualified for a Buckinghamshire grammar school."

#### FOI cases involving the objectors and TGSB and/or GL assessment

19. The objectors have provided me with judgements from four tribunal hearings in which one of the objectors has been an appellant against a decision of the Information Commissioner (IC). These are summarised, for background information, as follows:

- a. Appeal Number EA/2020/0310 in the First-tier Tribunal, with the IC as first respondent and TBGS as second respondent,

This case was heard on 17 August 2021, and 7 and 8 February 2022

The background is that various data had been requested in an FOI request, including a copy of the "detailed statistical analysis", as referred to in a letter from TGSB and GL assessment who wrote to the parents of children who sat the Secondary Transfer Test on 12 September 2019, explaining what steps had been taken in the marking of the test to account for some errors in the test papers. Other specific information was also requested, if not included.

TBGS did not supply most of the information, and the IC's Decision Notice (IC-47086-D9P3) shows that the IC did not require any steps to be taken.

This appeal was against this decision of the IC. It was dismissed.

- b. Case Number UA-2022-000677-GIA in the Upper Tribunal

This decision date for this case was 5 July 2023.

The decision of the First-tier Tribunal under reference EA/2020/0310 was found to have involved the making of an error in point of law.

The appeal was allowed, and the case was remitted to be reconsidered.

- c. Appeal Number EA/2020/0310 - reheard in the First-tier Tribunal

The decision date for this case was 18 March 2024.

The case was reheard, and the tribunal determined, by a majority, that the appeal against the decision of the IC should be dismissed.

- d. Case Reference: EA/2022/0245 in the First-tier Tribunal, with the IC as first respondent and The Lincolnshire Consortium of Grammar Schools as second respondent.

The decision date for this case was 26 July 2024.

The background is that the objector had submitted an FOI request to the Consortium, which included the following:

“The anonymised data for tests taken in 2019 for entry to the grammar schools in the Consortium in September 2020. For each candidate who sat the test the following information to be disclosed:

- Week of birth (the date of birth data to be aggregated to the nearest week)
- Verbal reasoning raw score
- Verbal reasoning standardised
- Non-verbal reasoning raw score
- Non-verbal reasoning standardised
- Total age weighted score

[...]”

The IC’s Decision Notice (IC-82675-K0J4) upheld the Consortium’s refusal to supply the above information.

The tribunal found that the Decision Notice was not in accordance with the law and issued a Substituted Decision Notice requiring the Consortium to disclose the above information.

## Consideration of Case

### Preliminary point – transparency and clarity

20. Central to the objectors’ arguments is their view that aspects of the arrangements are not transparent, to the extent that this renders the arrangements not compliant with paragraph 14 of the Code. The word “transparency” occurs on numerous occasions in the detailed case, and the following extract provides a sense of the objectors’ view:

“In order for parents to understand what is done, and to be confident about the ‘fairness’ of the overall process both the published admission arrangements, and the way in which they are implemented, must be sufficiently clear (**or ‘transparent’ which is used herein as an exact synonym** [my emphasis]) for such judgements to

be made. That transparency can only be achieved if sufficient information is volunteered by the admission authority or, at the very least, provided on request should the need arise.”

21. I note at this point that the word “transparent” only occurs three times in the Code, as follows:

- a. Paragraph 1.34: “Admission authorities that decide to use random allocation when schools are oversubscribed must set out clearly how this will operate, ensuring that arrangements are transparent...”
- b. Paragraph 1.15: “Admission authorities may wish to name a primary or middle school as a feeder school. The selection of a feeder school or schools as an oversubscription criterion must be transparent...”
- c. Paragraph 1.42: “Admission authorities may give priority in their oversubscription criteria to children eligible for the early years pupil premium, the pupil premium, or the service premium who: a) are in a nursery class which is part of the school; or b) attend a nursery that is established and run by the school. The nursery must be named in the admission arrangements and its selection must be transparent...”

None of these uses of “transparent” relate to the matters raised in the objection. There is no requirement in the Code for the testing arrangements for selective schools to be transparent. The relevant requirement is set out in paragraph 1.17 which states that the admission authorities for all selective schools must publish the entry requirements for a selective place and the process for such selection in their admission arrangements.

22. Paragraph 14 of the Code does indeed require that the practices and the criteria used to decide the allocation of school places are “clear”. However, the meaning of the word “clear” is expanded upon in this context. It is stated that “Parents should be able to look at a set of arrangements and understand easily how places for that school will be allocated”. I do not agree with the objectors that the requirement for clarity in paragraph 14 means that that every detail of every aspect of the arrangements has to be set out in full. A logical interpretation of paragraph 14 is that it requires that the arrangements are not vague, ambiguous, or confusing, so that they are easily understood by parents. The inclusion of a high level of detailed information, some of which may be difficult for some parents to understand, risks rendering the arrangements confusing and therefore less clear. In practice this means that when a set of arrangements is drafted, there is a trade-off between the level of detail necessary to be included and ease of comprehension.

23. I do not agree with the objectors that the word “transparent” can be substituted as a synonym for “clear” in my consideration of the compliance of the arrangements with paragraph 14 of the Code. The Code uses the words “clear” and “transparent” in different contexts for different purposes. I find that “clear” means unambiguous and easily understood while “transparent” means that the underlying reason or rationale should be apparent.

24. I conclude this section by observing that the objectors have left me in no doubt that they view aspects of TBGS in a negative light. For example, they say:

“...[there is] an endemic culture of secrecy that militates against the clarity required by the Code”

and

“...[there is] evidence that TBGS lacks the necessary expertise to organise satisfactory test arrangements”.

In relation to the way in which the schools have decided to process the scores so that the qualifying score is the same each year, one reason for which is to make matters clearer for parents, the objectors say:

“The concept that the qualifying score of 121 is fixed and immutable is false and misleading. Parents should be able to look at arrangements and understand how they work. How can they do that if the ‘fixed score’ doesn’t reflect an entirely ‘fixed standard’? Rather than using all means to maintain that fiction, TBGS should replace secrecy with openness.”

25. It is not the role of an adjudicator to rule on allegations about the culture and ethos of an organisation, nor is it in my jurisdiction to deal with matters about disclosure and non-disclosure of information which fall under the Freedom of Information Act (2000) and are under the jurisdiction of the Information Commissioner’s Office. Rather, I stress that in this determination I will only be considering whether the arrangements conform with the Code, and so will only be addressing those aspects of the objector’s views about “transparency” insofar as they relate to the Code, rather than as they relate to the issues discussed in the various tribunal rulings, such as public interest, commercial confidentiality, and openness and accountability.

#### Part 1 of the objection – does the reliability of the tests make them intrinsically unfair?

26. Key to the objectors’ arguments about the fairness of the testing process is the concept of “reliability” as it applies to assessment of ability through a testing process, and as a proxy measure for fairness. The concept, definition and means of measurement of the “reliability” of a test are all matters to which statisticians and educationalists have devoted many hours. As one of the tribunal rulings puts it:

“It is a subject which stirs strong passions, not only among statisticians. Fortunately for us, it is not our function to reach any view on the arguments...”

27. The objectors have drawn my attention to a paper entitled “Reliability, Validity and all that Jazz”, written in 2001 by Dylan Wiliam, which goes into considerable detail about the notion of the reliability (or otherwise) of various educational assessments.

28. In this paper, Wiliam says:

“If a student attempts a test several times, even if no learning takes place, the student will not get the same score each time—the student might not feel very ‘sharp’, the marker may be more or less generous, or the handwriting might be a little bit clearer so the marker can understand the answer. A further source of unreliability (usually the largest) concerns the particular choice of items. A test is constructed by choosing a set of items from a much bigger pool of potential items. Any particular set of items that are actually included will benefit some students (e.g. those that happen to have revised those topics recently) and not others.”

29. I note at this point that reliability measures are generally expressed on a scale from 0 to 1 (or as a percentage) where zero is unreliable and 1 (or 100%) would represent perfect reliability.

30. The objectors have also supplied me with extensive material on the nature of reliability which they consider relevant to the TBGS 11+ test and to 11+ testing in general. They have also responded in great detail to my questions about how different aspects of reliability can be measured from a dataset, and I note that a particular aspect of their argument relates to how aspects of the reliability of an individual’s score can be measured and reported. I will address these matters in more detail later, but at the moment I will consider the matter of “overall reliability”, as a single numerical measure (as opposed to any measure that applies to a particular individual).

31. The following statement from the objectors is important:

“...it is well known and understood that no academic test, however well-constructed, can be 100% reliable or consistent. There will always be some variation between the score achieved by a candidate on one day, and the same candidate taking a different example of the same test shortly afterwards. There are well established statistical techniques that allow (given sufficient real test data) the extent of this effect to be measured (Feldt, Steffen, & Gupta (1985), Coe et al (2008) and Cizek (2009).) This is known as “test/retest reliability” and can be expressed as a percentage. Based on the available research, our estimate is that overall 11+ assessment reliability can be expected to be about 85%.”

32. The objectors make the following argument as to the link between reliability and fairness:

“Any test is subject to a degree of measurement error and consequently this means that a 100% fair test is an impossibility. Because such testing is allowed by statute, it follows that something less than perfection must be permissible. Therefore, a judgement needs to be made, in the first instance by parents, but ultimately by the Adjudicator, as to how bad the arrangements can be allowed to be, before they fall below the standard of fairness required by the Code. In the case of any test that an admissions authority might choose to use, we would suggest that ‘test reliability’ is a good proxy for ‘fairness.’ Happily, there are highly objective statistical measures for test reliability which are available to decisionmakers as tools to determine the

fairness of any tests. At least, they are available for this purpose if sufficient data is provided to allow them to be applied.”

33. I agree with the objectors that the fairness of the TBGS arrangements is linked to the reliability of the tests. I also agree with their point that because testing for ability is lawful, the arrangements can still meet the requirements for fairness with a reliability measure of less than 100%.

34. I also agree that in general terms, there exist certain “objective statistical measures” that place a percentage value on reliability. However, I note here that there is no clear basis for deciding the verbal description to be allocated to that value, and the interpretation of a measure is likely to depend on what the observer considers to be a “good enough” value. For example, if an observer considers 78% as the boundary between “good enough” and “bad enough”, then a measure of 85% could be considered good enough for the tests to be considered reliable. But if the observer considers 88% to be the boundary, then the same measure of 85% will be considered unreliable. In other words, whilst an objective measure may produce a value, the interpretation of that value will be subjective.

35. I note here that William states that:

“The reliability of tests produced in schools is typically around 0.7 to 0.8, while that for commercially produced educational tests range from 0.8 to 0.9”

When I consider this statement alongside the estimate of overall reliability of 85% made by the objectors, I consider that I have little evidence in front of me to conclude that the reliability of the TBGS tests is out of line with what can be expected from a cognitive assessment test constructed by an external provider. I therefore do not determine that the tests are sufficiently intrinsically unfair, on the basis of their reliability, for them not to conform with the requirement of fairness in paragraph 14 of the Code. I do not uphold this part of the objection.

#### Part 2 of the objection - revised reporting proposals and the impact of reliability on test scores

36. The objectors raise the point that if results were communicated in a particular way, then parents and others would be able to understand the impact of reliability on their child’s score.

37. Their argument here is essentially that because the tests have a level of unreliability, there is an element of randomness to the actual score achieved by a child. It is therefore better to regard an individual’s standardised test score as a “signal” which is subject to “noise”, which can be modelled as a probability distribution rather than just a fixed value. In this model, the “actual score” achieved is at the mid-point of a bell curve which incorporates the “true value” somewhere, meaning that the “true value” could be higher or lower than the actual score. The “spread” of the bell curve depends on the reliability applicable to that child’s score; if the reliability is high then the spread is narrow and the “true value” is close to the actual value. Conversely if the reliability is low, the spread is much wider and there is

a higher possibility that true value is further away from the actual achieved score. A statistician would recognise this as describing a Normal distribution, whose mean is the actual score and whose standard deviation is some function of the reliability.

38. The objectors say that if a score is viewed in this way, then it is possible to communicate the result with some additional information which conveys the “noise” that applies to the “signal”. They suggest the following two methods for this:

- a. Each result could be accompanied by some sort of “confidence limit”, which expresses a range either side of the score accompanied by a percentage likelihood of the true score being within that range.
- b. Each result could be accompanied by a “misclassification probability”, in which results, particularly those close to the 121 threshold, are accompanied by a probability that the true score is the other side of the qualification boundary.

39. Confidence limits are frequently expressed as boundary values between which we can be reasonably sure the true value lies. For example, if a child has a score that is expressed as “120 with a 95% confidence limit of +/- 10”, it means that we can be 95% certain that the true value lies between 110 and 130. I make the point here that such limits are likely to be more useful in expressing the ranges where it is unlikely to find the true value, in this example below 110 or above 130. This is often how confidence limits are used in medicine, where a reading is considered normal within a fairly broad range, but where outlying results merit further investigation. Also, whilst a range of values lie within the confidence limit range, not all have the same likelihood. It is much more probable that the true value is closer to the actual value rather than at the extremities of the range. Taking these points into account, I am unpersuaded that the publication of a confidence limit will add much meaning to an individual score, and I note here that results from other high-stakes testing, such as GCSEs and A levels, are not accompanied by such limits.

40. The alternative method of showing the impact of reliability upon a score is demonstrated in the following diagram, which shows a typical bell curve for a child who has achieved a score (the mid-point) that is less than the qualifying score, marked as  $z$  on the diagram.



In this curve, the shaded area under the curve shows the probability of correct classification, whilst the unshaded area shows the misclassification probability, being the region where the true score is higher than the  $z$ , the qualifying score. The objectors suggest

that parents should receive the value of their child's misclassification probability, alongside the test result.

41. At this point, it is important to note that when TBGS produce a final standardised and weighted score, it is rounded down to the nearest integer. This is a crucial detail in my consideration. The effect of this rounding is that any child with a reported score of 120 or lower will have scored less than 121 (the qualifying score) before any rounding. This means that the corresponding diagram for them will look like the one above – the qualifying score,  $z$ , will be 121, and their score will be the midpoint (below 121). The diagram shows that the chance of misclassification will be less than 50%. In other words, the civil standard of proof will have been met for deeming all such students as not qualified. It is of course possible that a student with a score of 120 could have a true score that is 121 or higher, but it is more likely than not that the true score is below 121. There is one more crucial detail here - this conclusion is reached regardless of the spread of the bell curve. As I have explained above, the spread is related to the reliability, and so it is clear that the conclusion (namely that the civil standard of proof has been met in deeming a student with a score of 120 as not qualified) is independent of any reliability measure.

42. The outcome of this line of reasoning is that I do not see how the reporting of results with additional information in either of the ways the objectors propose contributes anything to the clarity or fairness of the arrangements in relation to the award of qualifying status via testing. The arguments I have set out above show that regardless of reliability, confidence limits or misclassification probabilities, the only reasonable course of action open to TBGS is to consider that "the best estimate of the score is the score" and therefore to deem initially as unqualified those who do not achieve a score of 121 or better.

43. I have additionally considered whether the publication of confidence limits and / or misclassification probabilities might assist in the Selection Review (SR) process. On the face of it, the objectors' arguments might seem to have some merit, in that the SR panel would be able to look at any score within a few marks of the threshold and consider the probability that the true score was in fact above the threshold (I have already explained that these probabilities will all be below 50% and it is always the case that it is more likely than not that the child is correctly classified). However, I have reproduced the details of the SR process at length earlier in this determination, and it is clear from those details that this process allows a great deal of information about a child to be put forward, to support a parent in advocating for a change in their child's qualification status. It is my considered view that the only additional conclusion that could be drawn by an SR panel from a misclassification probability is the (rather obvious) conclusion that within a group of children who have not achieved a qualifying score, those with the lower scores are even more likely to have been properly classified than those whose scores are closer to the qualifying score.

44. SR panels make evidence-based decisions to award qualified status and there is extensive data on the TBGS website that summarises the outcomes of SR panels. This data shows a general trend that the closer a score is to 121, the more likely it is that the SR process will change a child's status will be changed to qualified. The SR panel decisions are therefore consistent with what is to be expected, and I am not persuaded that the



addition of information conveyed by a misclassification probability is likely to contribute to the process in any significant or meaningful way.

45. Taking into account all the above arguments, I am not persuaded that the objectors' proposals for revised reporting would add anything to the clarity or fairness of the arrangements as they currently stand. I also have no reason to determine that they are required by the Code. I therefore do not uphold this part of the objection. Furthermore, even if I were to determine that the testing procedure operated unfairly, I would not have the power to determine that specific revisions be made to the arrangements as this would be beyond my jurisdiction.

46. I stress here that I do not dispute the objectors' analysis of the test score as a probability distribution. In any assessment of academic ability, there will be an inherent degree of unreliability that could be expressed in the ways the objectors suggest. TBGS tell me that they have "at no time asserted that it is impossible for confidence intervals to be produced for the STT", and I do not dispute the general principle that any numerical measure, be it an academic assessment or otherwise, can only be completely understood if there is more information about measurement error. However, I refer the reader to the earlier section on the difference between transparency and clarity, and I emphasise the point I have made earlier that paragraph 14 of the Code, which sets the legal parameters in respect of the practices and criteria used to allocate places, requires that parents should easily understand how places are allocated, and that is initially on the basis of the actual score rather than any other measure.

### Part 3 of the objection – proposed improvements to the testing process

47. The objection stated that:

"We also acknowledge that there is more than one way in which the process could be improved. For example, administering tests twice to each candidate with the better performance being used to decide eligibility. Thereby doubling the chance that any candidate will fall at the 'right' end of the zone of uncertainty on at least one test."

48. I queried the mathematical basis for this assertion with the objectors, since it is the case that if, for example, the chance of success at a task is 0.5, then repeating the task increases the chance of at least one success to 0.75 (calculated as  $1 - 0.5^2$ ), which is not the same as doubling 0.5. The objectors' response was:

"This was not a mathematically based assertion. It was simply a statement of the obvious that 'two bites at the cherry' is [sic], colloquially, 'twice as good as one.'"

49. I do not dispute that if the testing process were to be doubled up, it is entirely possible that some candidates would achieve a qualifying score on one test but not on the other. However, I make the following observations:

a. The objectors say that:

“The advantage of the ‘best score counts’ is that it reduces the impact of “extrinsic bad day” failures. In other words, it allows candidates who under-perform for personal reasons on one day to be judged solely on their better performance, rather than having their average dragged down by the other one.”

However, this is predicated on an assumption that the lower score is the result of a “bad day”, rather than on the equally possible scenario that the higher score is due to a “lucky day” and consequently that the lower score is actually a truer reflection of ability. In short, it is not clear to me why “best of two” is any more logical or rational than “worst of two”.

- b. If the scores from the two occasions are very different, it is not possible to determine whether one of them is an outlier and if so which one, or whether both are deviations from a true value that sits between them. The only way to be clearer is to take more tests.
- c. If each set of tests were standardised and scaled, and the best score taken, then a further scaling process would be necessary for TBGS to ensure that there is still a link between the number of children awarded a score of 121 or better and the number of places available in TBGS schools. Hence the overall ratio of number of qualifying children to places available would still be broadly the same.
- d. It is undoubtedly the case that if results from two occasions were averaged, then there would be an increase in reliability. Effectively the test would have been doubled in length, which improves reliability, and the “good day or bad day” effect would be mitigated. However, as explained in point b above, two occasions are less reliable than three, three are less than reliable than four, and so on.

50. Having considered the above points, I conclude that ultimately it is for an admission authority to decide whether or not to test on multiple occasions. In the case of TBGS I have already stated that I view the overall reliability of the testing process, as currently determined, to be at an acceptable level. Whilst it is open to TBGS to increase the length of their tests and/or conduct them on multiple occasions, I am not persuaded that these steps are necessary in order for the arrangements to comply with paragraph 14 of the Code. I therefore do not uphold this part of the objection. In any event, as I have already stated, even if I were to determine that the testing procedure operated unfairly, I would not have the power to determine that specific revisions be made to the arrangements as this would be beyond my jurisdiction. It would certainly be beyond my jurisdiction to require the admission authority to arrange for each applicant to be tested on multiple occasions.

#### Part 4 – the remaining proposed additions to the arrangements.

51. In this section, I will address other matters that relate to what the objectors say needs to be added to the arrangements in order to conform with the Code, such as full details of the methodology used to turn raw answer data into the binary 11+ status, the age-standardisation and scaling algorithms, and a commitment to releasing a complete

anonymised data set on request to any third party who wishes to conduct their own analysis.

52. At the outset of this section, I note that the way in which a set of admission arrangements is implemented lies outside my jurisdiction. I make this point because the objectors list some recent situations involving selective schools in which they say shows that there have been “unfair practices and outcomes”, including some data processing errors. They also say:

“... there is sufficient circumstantial evidence ... to suggest that there is a significant risk that the nature of the tests TBGS (on behalf of the individual admissions authorities) is intending to use, the way they are marked and the ways in which the ‘raw marks’ outcomes of tests are translated (via standardisation and/or other algorithms and adjustments) into a pass/fail decision may be procedurally unfair when they take place.”

I observe here that alleged past procedural errors are not in my jurisdiction and so I make no further comment on the above statement.

53. There is also no requirement in the Code for arrangements to set out in any detail the procedures by which children’s data is processed, nor the methods by which such processes are checked for accuracy. Neither are admission authorities obliged by the Code to release the data to third parties for the purpose of checking, for example, that age-weighted standardised scores have been correctly calculated. That is not to say that verification and checking of data is an unimportant matter, merely that it is not covered by the Code and is not in my jurisdiction, and therefore is not a matter on which I can rule.

54. I also do not consider that the Code requires the publication of factors and algorithms relating to age-standardisation and scaling for any other reason. That is because in my view, such information is unlikely to add any clarity to the arrangements for the ordinary parent. Indeed, any such detail may obscure the key elements which all parents need to know. In my view those key elements, as outlined in an early section of this determination, are perfectly clear. I also note that some of this information can only be issued retrospectively due to the nature of the standardisation that operates within TBGS. It is difficult to see how the inclusion of previous years’ factors would have added anything to the clarity and fairness of the arrangements for admission in September 2025.

55. It is of course open to anyone to make an objection that a particular aspect of a set of arrangements is unfair. Had an objection to the fairness of age-standardisation been made, then that would have been investigated, but I note that it is perfectly possible to investigate the effect of the standardisation on different ages without disclosure of the actual factors themselves.

56. I now turn to the wider issue of reliability. In one of their responses to my queries, the objectors say that:

“parents should be given the outcome of a recognised ‘test/retest reliability’ measure calculated for that iteration of the test, alongside the results for each child.”

I have already addressed the suggestions for the ways in which the impact of reliability could be individually communicated for each child, for example by confidence limits or misclassification probabilities, but in this section I will address the issues that arise from the consideration of wider measures of test reliability.

57. I have extensive responses from the objectors in response to my queries about some of the technical matters underpinning their objection. On 11 September I wrote to them to clarify my understanding of some of the points they had made, including the interpretation of an academic paper they had cited, together with some source material which I had researched which I considered relevant, including two documents published by Ofqual<sup>1</sup> and some websites<sup>2</sup>. The combined responses to these queries amounted to almost 12000 words. I do not intend to go into great detail about this correspondence, but I have considered it alongside responses from TBGS, the original objection documents, and the source material shared with the objectors. Having done so, I make the following observations:

a. The “reliability” of a test is a multi-faceted concept – it is not the same as “validity” (which in this case would be a measure of the extent to which the TBGS 11+ tests are a “true measure of ability”, as required by paragraph 1.31 of the Code)

b. Aspects of reliability relevant to the 11+ include:

“internal consistency reliability” (which the objectors define in one document as “the extent to which any test or examination actually tests what it is supposed to test and not anything else” and in another document as “a measure of the homogeneity of the test, i.e. how much each individual question assesses the same competencies as the rest of the assessment.”)

and

“occasion reliability” (or “test/retest reliability”) which they describe as “a measure of how consistent a candidate’s score is if the test administered multiple times.”

c. It appears that it is not possible to quantify the “having a good or bad day effect” from data analysis and therefore any published measures of overall reliability may not truly reflect the impact of this.

---

<sup>1</sup> “Conceptualising and interpreting reliability” (Ofqual 2009), “Estimates of Reliability of Qualifications” (Ofqual 2010)

<sup>2</sup> <https://www.statology.org/split-half-reliability/>, <https://researchmethod.net/split-half-reliability/>, <https://www.gov.uk/government/publications/reliability-of-assessment-compendium/introduction-to-the-concept-of-reliability>.

- d. It is possible for a body to analyse a dataset from a test, generate a numerical statistical measure that relates to some aspect of the test's reliability, and then issue a statement. For example, TBGS tell me that:

“GL Assessment uses Cronbach's Alpha to measure internal reliability. This formula produces values ranging from 0 to 1. Values of 0.8 and above are considered to be very good. TBGS do not publish reliability statistics for the Buckinghamshire STT. However, we can confirm that all tests administered as part of GL Assessment's contract with TBGS have internal reliability scores well in excess of 0.8, both for the overall weighted score and for the individual component scores.”

- e. However, it is clear from the objectors' correspondence that whenever a body issues a statement such as this, they open themselves up to questions such as what aspect of reliability was being measured, whether the statistical measure used was appropriate in that situation, what interpretation should be placed on the numbers that are produced, and indeed whether it is appropriate to use a reliability measure that relates to the cohort in general rather than to express the reliability associated with an individual's score. I consider that any third party who was provided with data for analysis would be likely to have their conclusions similarly questioned.

- f. The objectors say that:

“The objection raised with the Adjudicator is predicated on the Admissions Code that states, “**Parents** should be able to look at a set of arrangements and understand easily how places for that school will be allocated.” **The objective must therefore include a requirement for reliability to be clearly explained in non-technical language.**[my emphasis]”

When I shared with the objectors, for their comment, the documents and websites that I had researched and which I considered might assist in explaining the various aspects of reliability outlined in point b above, I did so in relation to technical questions I had about the some of the objectors' assertions about measuring occasion reliability. Two of these websites also contained what I considered to be language that would be accessible to an ordinary intelligent individual. Alongside detailed answers to my questions, I also received comments from one of the objectors, describing one of these websites as “Enid Blyton's Noddy does Statistics” and the other as “Mickey Mouse”.

58. I consider that the observations in the previous paragraph demonstrate the complex, technical and sometimes contentious nature of the matter of measurement of test reliability. I am therefore not persuaded that the arrangements would be any clearer if they contained “an independently verified test/retest reliability index”, nor that they are unclear because they currently do not.

59. Taking into account all the above, I am not persuaded that the proposed additions add anything to the clarity or fairness of the arrangements as they currently stand. Moreover, I consider that the proposed additions have the potential to lead to confusion amongst parents and obscure the key points about the arrangements that they need to know and understand. I also have no reason to determine that they are required by the Code. I therefore do not uphold this part of the objection.

## Other Matters

60. In 10 of the 13 sets of arrangements, it appeared to me that the description of the “Selection Review” process might not be sufficiently clear to comply with the requirements of paragraph 14 of the Code.

61. Three of the admission authorities contain within their arrangements a link to a Buckinghamshire County Council website that describes the SR process in extensive detail. These are the arrangements for Beaconsfield High School, for Chesham Grammar School, and for Sir Henry Floyd Grammar School. The arrangements for the other 10 admission authorities refer to little more than the existence of the process. TBGS have confirmed that the SR process is in fact the same for those other authorities.

62. It is my view that details of the SR process are an essential part of the arrangements since it is those details which explain the basis on which qualified status can be awarded, together with steps taken to ensure that the process is fair, clear and objective. Therefore, I consider that any set of arrangements without either details, or a link to those details, is unclear and does not conform with paragraph 14 of the Code.

63. This matter is one which can be easily and quickly addressed by the admission authorities to which it applies. Given the timetable for the secondary admissions process, I require that the relevant sets of arrangements for those 10 TBGS schools who are not mentioned earlier are revised within two weeks of the date of this determination.

## Determination

In accordance with section 88H(4) of the School Standards and Framework Act 1998, I do not uphold the objections to the admission arrangements for September 2025 determined by the admission authorities for the 13 grammar schools in Buckinghamshire.

I have also considered the various arrangements in accordance with section 88I(5) and find there are other matters which do not conform with the requirements relating to some of the admission arrangements in the ways set out in this determination.

By virtue of section 88K(2) the adjudicator's decision is binding on the relevant admission authorities. The School Admissions Code requires those admission authorities to revise their admission arrangements within the time scales set out in this determination.

Dated: 16 October 2024

Signed:

Schools Adjudicator: Clive Sentance