

English Bridge Union



National Grading Scheme

Full description

Version 14

1 December 2014

Authors

The NGS Working Group:

Present and past members of the Working Group include:

Mike Christie (Project Manager and Programmer)

Barrie Partridge

Paul Bond

Paul Habershon

Andy Kittridge

Peter Lee

Anthony Moon

John Probst

Jerry Rudge

The late Dr John Carter

Barry Capal (EBU General Manager)

Michael Clark (EBU IT Support Officer and web master)

This Guide

This guide provides full details of the National Grading Scheme itself, but please see also on the EBU web site a guide for scorers “Procedural Guide for scorers”

New for Version 14: Information on Teams events, which are now being graded, material from a feature (NGS Myths by Michael Clark) in English Bridge August 2014, and a research paper by Mike Christie (see FR18).

Contents

Page

Introduction to the National Grading Scheme

Objectives of the National Grading Scheme	5
NGS Grades	5
Calculating your grade	6
Which events are graded?	7
Viewing your grade	7
Accuracy of grades in the NGS	7
Partnership grades	8

Full Details of the scheme

Which events count toward the scheme?	9
Decay of effect of previous sessions	9
Strength of opponents and strength of field	11
Calculating Session Grades for pairs events	12
1. Match pointed pairs	12
2. Butler pairs	13
3. Cross-IMPed pairs	13
4. Aggregate pairs	14
Calculating event grades for teams of four events	14
Grades for partnerships	15
Grade initialization	16

Diffusion	17
Assisting diffusion	18
Limitations to diffusion	19
Chronology	19
Using the national grading scheme for club handicapped events	20
Further information on the scheme	20

Frequently Asked Questions

1. What does my EBU numeric grade mean?	21
2. How are grade bands and numeric bands related?	21
3. How are EBU grades calculated?	21
4. How many sessions do I have to play to get a grade?	22
5. Will my grade change by the same amount as that of my partner after a session together?	22
6. How does the effect of a board or a session that I play today diminish over a period of time?	22
7. How can I work out, as I start to play an evening duplicate at my club, what % Match Points score I need to achieve in order to maintain my grade at its present level?	23
8. Is there not an advantage for regular partnerships?	23
9. Can I opt out of the scheme?	24
10. Can a "host" player have an event excluded from the scheme?	24
11. Why do the grades seem to be so sensitive, and why is so much weighting given to the session just played?	24
12. Looking at my sessions on the website, why does my grade sometimes seem go down after a good result?	24
13. I was partnered with a "King" graded player last night, but he/she didn't seem to play at a level that I would have expected of a "King"	25
14. Why should my grade suffer when partner plays like an idiot?	25
15. Do I need to play with different partners to have an accurate grade?	25
16. Why does my grade seem to rise in small amounts and fall in large amounts?	26
17. How often will grades be updated?	26
18. Why hasn't one of my recent sessions shown up in my NGS data?	26
19. What about Simultaneous Pairs and other multi-section events?	26
20. How are grades standardised across the country?	27
21. Should you really be trying to compare IMPs with MPs? In any case, I play better at one than the other	27
22. I perform better in a strong field / in a weak field / with a stronger partner / with a weaker partner. Can I beat the system?	27
23. Do other countries have similar grading schemes?	28
24. What about when I have a sit-out?	28
25. What happens when I am given an "average" on a board?	28
26. Why do many of my recorded sessions indicate that I played 24 boards when I know I played more or fewer boards?	28
27. When will Teams results be included in the NGS?	28
28. I mostly play with my regular partner, Ann, and expected to have almost the same grade. Why is my grade so much less than hers? Surely there's a mistake?	29
29. Non-graded sessions	30
30. Teaching sessions.	30
31. Student and "Pro-Am" sessions	30
32. Novice sessions	31
33. How is diffusion progressing?	31
34. Stratification and NGS	31
35. Converting NGS to Handicaps	32

36. How can we set up handicapped events if we have club members opting for privacy?	32
37. "The NGS has meant that I won't play with weak partners"	32
38. Are there plans to export the NGS? Scotland, Wales, further...	33
39. Can we improve the display of the NGS on the web site?	33
40. How often does the NGS re-calculate and when?	33
41. Are clubs generally processing their P2P submissions promptly?	33
42. Have you had the "gain" turned up well above 1? Has that made it more likely that there is under or over shoot?	33
43. Average grade of partnership and partnership grade	33
44. Initialisation	34
45. My partner and I used to play occasionally with other people but for the last 1000+ boards we have only played with each other. Why are our NGS grades different?	34

Further Reading

1. Is the average grade exactly 50%?	36
2. What is the difference between "straight line" decay and "exponential" decay, and why have we opted for "straight line" decay?	36
3. Can a single board have a greater impact on my grade if some other scoring than Match Points is being used?	37
4. Would it be reasonable to put a limit on the influence of an exceptionally good or bad session?	38
5. Can I use the information from the National Grading Scheme to work out the probability of how I will score this evening at my local club?	38
6. What are the arguments for and against introducing additional weighting for certain events?	39
7. How might we handicap team events using the National Grading Scheme data?	39
8. What types of Pairs events are we not considering for the National Grading Scheme?	40
9. How did you get the formula for Aggregate Pairs?	40
10. What happens if I have been an EBU member for the last three years but played fewer than 300 boards in that time?	42
11. What happens if I stop playing bridge for three years and then restart?	42
12. How does NGS cope with unknown players in a session?	42
13. The "Aces and Eights" issue.	43
14. Volatility and Decay.	45
15. Linearity of results.	45
16. Other interesting comments from the Internet forums	46
17. Concerns over grading Teams events.	47
18. Research arising from the NGS – Variability in Duplicate Pairs Bridge Events	48



Introduction to the National Grading Scheme

This introduction states the objectives of the NGS and explains how it affects you, a playing member of the EBU.

Objectives of the National Grading Scheme

The NGS is a measure of performance for EBU members, which is relevant to the vast majority of the membership, from novices to ordinary club players to Internationals.

Hence the objectives for the NGS are:

- To provide a fair and trustworthy measure of an individual's current performance when playing duplicate bridge, which is easy to understand and reflects changes (both up and down) in an individual's current standard of play.
- To enable new competitions to be constructed for use in Club, County and National competitions based on the current performance of individuals, through treating the measure of current performance as a "handicap".

It should be noted that the NGS is complementary to the Master Point Scheme, whose existing arrangement is unaltered. The two schemes are independent. The Master Point Scheme provides an indicator of an individual's aggregated lifetime achievement in bridge, while the Gold Point scheme provides a ranking among the top players who frequently play in National events, based on the last five years' performance. In contrast, the NGS indicates an individual's current playing ability based on the last 80 or so playing sessions at whatever level they play.

NGS Grades

The NGS calculates a current grade and a grade band for each EBU member, which is updated whenever new results for that player are received.

The value of your current grade is the scheme's estimate of the percentage score that you would achieve on average if partnering another player with the same current grade at Match Pointed Pairs in a field of nationally average strength.

An average strength player therefore has a grade of about 50%, but there are no limits to the values of grades. The strongest players have current grades of over 60% and the weakest under 40%. The spread of grade values has a bell-shaped distribution. (For the statisticians, it's approximately Normal with a standard deviation of around 5 to 6%.)

EBU grades have been divided into 13 playing card bands, ranging from 'Ace' at the top to 'Two' at the beginners' end. Each band covers a 2% range of grade values, with 'Eight' having a range of 49-51%.



Full Description of the National Grading Scheme

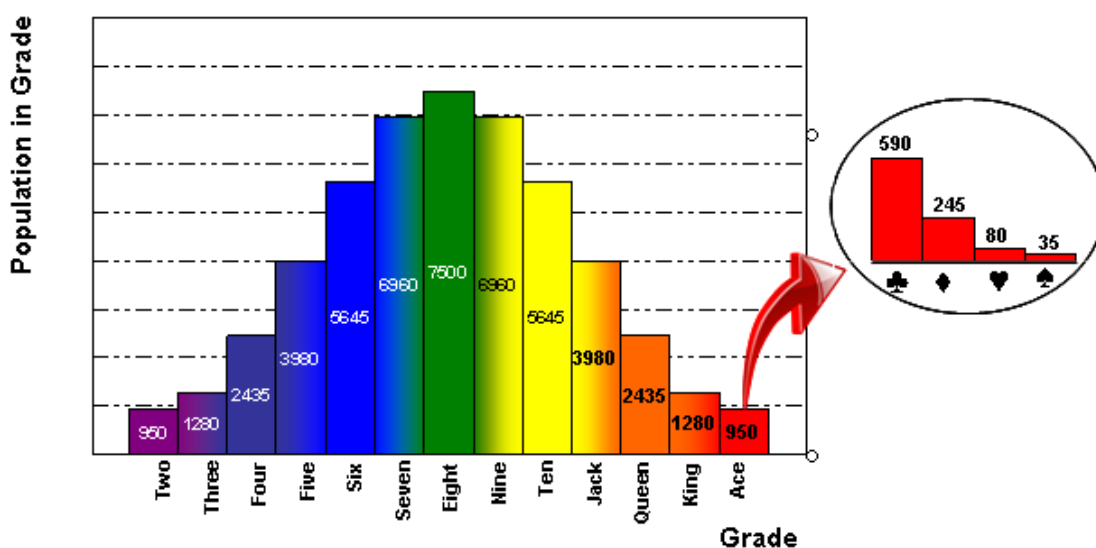
The 'Ace' band is subdivided into four further bands:

	% grading range
Ace of Spades	67+
Ace of Hearts	65-67
Ace of Diamonds	63-65
Ace of Clubs	61-63

The chart below shows how before the start of the scheme the spread of grades was expected to look when the scheme became mature, and based upon an underestimated 50,000 members. The current spread of grades may be seen at

<https://www.ebu.co.uk/ngs/stats>

While the spread is close to that expected, the average grade has settled at just below 49%.



Calculating your grade

Updating a current grade based on the scores you have achieved in duplicate sessions is not new. Similar methods have been tried and tested for more than 25 years in England, for example at the Young Chelsea and Sheffield Bridge Clubs and at the online club Bridge Club Live. However the details of the NGS scheme are unique to the EBU. The scheme commenced at the start of "Pay to Play" in April 2010.

For most events, the calculation is based on your overall result for the session rather than each individual board, though for some events we need to use the results of your boards versus specific opponents. When you play in an event your expected score is based on the average of your and your partner's current grades and on the average strength of the players against whom you are competing. The amount by which you over (or under) achieve this score is used to determine your 'session grade' for that session.

For events that aren't scored using Match Points, we convert the scores to something equivalent in Match Points percentages.

We use your session grades for all the events that comprise your most recent 2000 boards and take a weighted average of these to produce your updated current grade. The weighting of older sessions declines linearly and



contributes nothing to your current grade when you have played over 2000 boards more recently.

Until you have played 300 boards in graded sessions, your computed grade is still significantly affected by the initialisation process and is not published, though we still use the evolving grade in grade calculations. Only once you've played over 1000 boards within the last three years is your grade said to be "mature".

Which events are graded?

Currently, the NGS includes:

- All duplicate Pairs events (including Match Point and IMP scored).
- All Swiss Pairs events with Match Points or IMPs, whether or not they are converted to Victory Points.
- Multiple Teams and Swiss Teams with IMP scoring (whether or not the IMPs are converted to VPs)
- Teams league events where boards are played at more than two tables

It is not intended to include head to head matches between two teams of four.

The inclusion of events is subject to the bridge scoring program used.

Viewing your grade

At any time you can log in to the EBU website and see your current grade, your partnership details, and all the contributing session grades (so you can check when that disastrous session with Fred will no longer contribute!).

We also publish on the EBU website regularly updated grade lists of the top players, and grade lists of all players by County, which enable you to check a player's current grade. Grade lists by clubs are also available to club administrators.

Those who want to opt-out of the scheme can choose that their grade is not published. However, in order to maintain the integrity of the NGS, grade values are calculated and updated for all graded sessions and all EBU members since their grades affect the grades of the players against whom they play.

There's one exception to this. If a player volunteers as a "host" or "mentoring" partner for a club duplicate session they can decide in advance that their score, or their and their partner's score for that session should be ignored by the NGS. This is limited to one player or pair per session and is at the discretion of the organiser of that session.

Accuracy of grades in the NGS

Your current grade is subject to random fluctuations depending on your luck, just as a single session is. We expect the standard deviation of the error in your current grade to be around 2%, provided you have a typical mix of partners. This depends in part on the exact parameters used within the NGS.



Of course, if you and your partner only ever partner each other, the difference in your strengths will not be known to the NGS and you'll have the same current grade. There are other, uncommon, partnership patterns which do not allow the difference between an individual's grade and those of their partners to be accurately estimated. We continue to analyse data on grade variability as NGS data is collected.

The NGS is a **national** scheme, and the mixing between clubs and national or local tournaments has allowed differences between the average playing strength of different clubs to be reflected in their players' current grades. Most of this diffusion took place very quickly after the start of the scheme, but where there are clubs whose members rarely or never play in other graded sessions, full diffusion took some time longer, and we are now confident that members' grades are truly comparable with the national pool of EBU members.

Partnership grades

For each partnership you play in, the NGS will also calculate a partnership grade, based on between 300 and 2000 boards played with that partner in the last three years. Some players may like to track this as well as their individual grade, as a "mature" partnership grade is less susceptible to the random errors described above. However, unless you play a lot of bridge with several partners, comparing your grades with different partners is as likely to reflect the effects of chance as to show how well you gel with each partner.



Full Details of the scheme

This describes all the technical details of the scheme and the rationale behind these details. Some discussion is contained in the Frequently Asked Questions (FAQs) and Further Reading (FR) which are included at the end of the document.

The National Grading Scheme has as its foundation the work of a couple of engineers who have run schemes at Young Chelsea and Sheffield Bridge Clubs and the online club, Bridge Club Live, and who have had extensive observation and experience of what works. Significant additional input has been undertaken by some experienced mathematicians to ensure that the scheme is robust, and various tests have been undertaken.

Which events count toward the scheme?

The following seven common types of events count towards the scheme:

- Match Pointed Pairs
- Butler Pairs
- Cross-IMP Pairs
- Aggregate Pairs
- Swiss Pairs with MP or IMP scoring
- Multiple Teams with IMP scoring
- Swiss Teams with IMP scoring
- Some Team League events where boards are played at more than two tables

There are at least three other types of Pairs events that we are not including. These are much rarer and are discussed in the Further Reading section (FR 8). In addition, we will not be including Individual events as they are fairly rare and many of these are actually held more as social occasions.

We now include many types of Teams events as some scoring programs are now able to transmit the necessary data. We have no plans to include head to head Teams-of-Four matches with boards not being played in any other match, nor do we intend to include "Point-a-board" Teams events.

Decay of effect of previous sessions

We take account of the most recent 2000 boards played by a player (or by a partnership for a partnership grade), subject to a time limit of three years previously. That's roughly the last 80 sessions in club evening duplicate terms.

Before considering how the grade for a session is estimated, let us look at the details of the weighting. Let y refer to the session, with $y = 1$ the most recent, $y = 2$ the next most recent and so on, until $y = k$ refers to the earliest session considered in the last 2000 boards. Let x_y be the number of boards played in that session, with x_k if necessary reduced so that $x_1 + x_2 + \dots + x_k$ equals exactly 2000. Let g_y be the 'Session Grade' (SG – the calculation for this is discussed later) for that session. The weight attached to session y , w_y , is calculated by the formula:

$$w_1 = x_1 (2000 - x_1/2)$$

$$w_2 = x_2 (2000 - x_1 - x_2/2)$$

$$w_3 = x_3 (2000 - x_1 - x_2 - x_3/2)$$



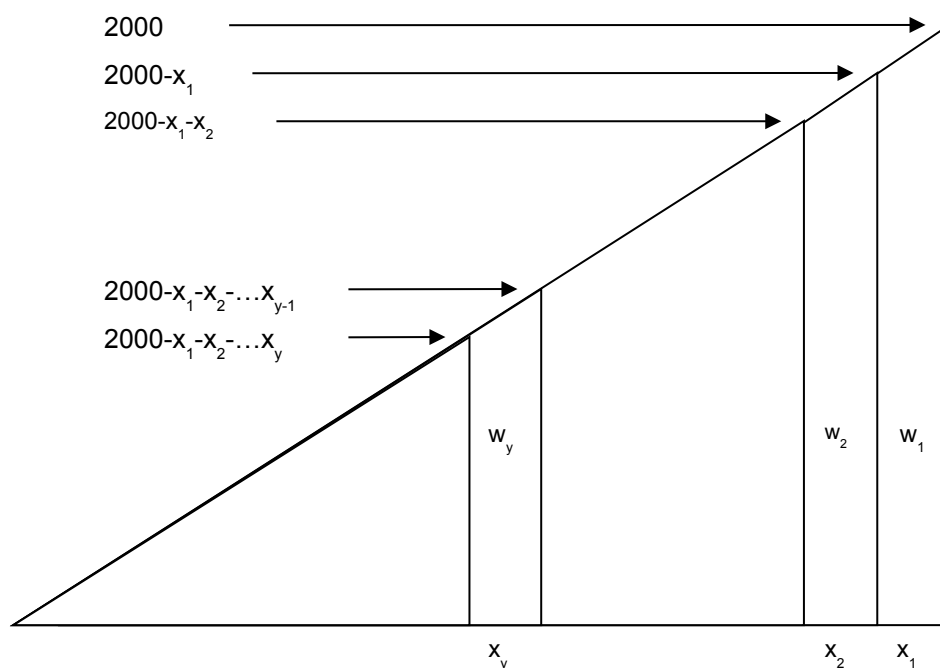
Full Description of the National Grading Scheme

$$w_y = x_y (2000 - x_1 - x_2 - \dots - x_{y-1} - x_y/2)$$

The player's current grade (CG) is then estimated by the formula:

$$CG = \frac{g_1 w_1 + g_2 w_2 + \dots + g_k w_k}{w_1 + w_2 + \dots + w_k}$$

The decay of the weighting factor for older events is shown diagrammatically below, where the most recent event is shown on the right:



w_y is the area indicated, with the width equal to x_y and the average height equal to $2000 - x_1 - x_2 - \dots - x_{y-1} - x_y/2$.

As the area of a triangle is half the base times the height, the sum of the weights is equal to $(2000 \times 2000) / 2 = 2,000,000$.

What happens if a player has played fewer than 2,000 boards, as will be true for everyone at the start of the scheme?

Where a player has played at least 300 boards but fewer than 2000 boards during the past three years, the player's grade is calculated based on the number of boards actually played during the past three years. The rate of decay remains the same, which means, in terms of the above triangle, that the bottom left portion is omitted.

Each board played in a given session has equal weighting. All this means that the straight line of decay looks in practice more like a series of small steps, but that doesn't matter, as we have taken into account the duration of each session.

Events which have multiple sessions, in which the players are divided into different sections in different sessions, produce a 'session grade' for each such session. Where a player plays in two or more sessions or events on the same day, the session grades are given equal weighting. Where a single event spans two or three days, the date given in the data sent to EBU HQ is deemed to be the date of play for grading purposes.



So, we are using a player's last 80ish sessions to work out some kind of weighted average of the player's session grades over each of those sessions.

Until April 2013, three years after the start of the scheme, we used each player's last 1000 boards rather than 2000 boards. This aided diffusion but increased the volatility of each player's grade.

This "straight line" method of decay has been chosen rather than the method used in most grading schemes so that session grades over 2000 boards ago have zero weight. This is discussed further in the Further Reading section (FR 2).

Strength of opponents and strength of field

The strength of a field is the average current grade of all the players in an event at the start of play of that event, and this can be used as a measure to compare the standard of an event with events played elsewhere across the country. Some clubs have a stronger "Strength of Field" than others and clubs that have more than one regular session each week may have significantly different Strengths of Field from one series to another. Also, National (and County) events tend to have much higher Strengths of Field than most club events.

When working out grades, we need to take into account the strength of opponents in order to judge what we have achieved in our results when we play. The remainder of this section comprises a discussion of how we go about this.

We need to establish a "Strength Of Opponents" factor, which we will abbreviate to "SOpp" for different types of event.

Let's just consider Pairs events before we move on to Teams events.

For the Bridge Club Live scheme, with Match Pointed Pairs play being mostly fairly casually organised in a very large field where you usually play few of the available boards and in variable directions, it is ideal to analyse boards individually using the strength of your opponents at the table.

For more controlled face to face (f2f, as they say online!) movements it is usually better to take a SOpp factor as being generally representative of the strength of the opponents. Normally it will make very little difference.

Take a 5 table full Howell. You play three boards against all nine of your opponent pairs. The SOpp factor will be the same as taking the average of all the opponents' current grades. Were it not for the fact that we exclude your own and your partner's current grade, the SOpp factor would be the same for everyone.

For a 2-winner Mitchell, we need a subtle variation. Let's say that you are playing a full table full movement Mitchell without arrow-switch, and you and all the other strong pairs are North South. You might think of taking a SOpp factor from the East West pairs, but think again. Our real opponents are those sharing the North South seats, as we are competing against them in the final ranking list. If we used the East West pairs to calculate SOpp, then overall the North South pairs will suffer a decline in grade, and the East West pairs will enjoy a rise, which wouldn't be fair.

Things are different for a Swiss Pairs Movement. Here we take each match as a separate stanza within the event as a whole. For each match, we are playing against a single pair of opponents. Here, as in Bridge Club Live, it is by far



best to take the grades of just your two opponents. The SOpp factor for one match is just the average current grade of your two opponents.

Your SOpp for the event as a whole is the average current grade of your opponents in each of your matches.

So the SOpp factor varies depending on type of movement:

For Swiss Events, we take the average current grade of your opponents for each match.

For 1 Winner Movements, we take the average current grade of all the other pairs.

For 2 Winner Movements, we take the average current grade of all the other pairs sitting in the same direction as you.

For Teams events, your SOpp is the average current grade of your opponents in each of your matches. For Multiple Teams events, where matches are short, we take the average current grade of all the other pairs sitting in the same direction as you. There is likely to be very little difference in method for Multiple Teams events as teams tend to comprise pairs of similar standard and the strength of North-South fields tends to be very close to the strength of East-West fields.

Calculating Session Grades for pairs events

1. Match pointed pairs

For Match Pointed Pairs, the formula for the expected performance for yourself and your partner for a session (or for a single match in a Swiss Pairs event) is

$$50 - \text{SOpp} + \text{the average of our and our partner's current grade (G)}$$

Let's refer to you as "a" and your partner as "b", so

$$\text{Expected performance} = 50 - \text{SOpp} + \frac{1}{2} G_a + \frac{1}{2} G_b$$

Suppose, you play a session and you score p%. Your score above your expected performance is then:

$$p - 50 + \text{SOpp} - \frac{1}{2} G_a - \frac{1}{2} G_b$$

From this we can set your session grade above your current grade by an amount proportional to the excess in your performance.

So we can set our Session Grade for the session that we have just played to

$$\text{SG}_a = G_a + \text{gain} \cdot (p - 50 + \text{SOpp} - \frac{1}{2} G_a - \frac{1}{2} G_b)$$

And equally, for our partner

$$\text{SG}_b = G_b + \text{gain} \cdot (p - 50 + \text{SOpp} - \frac{1}{2} G_a - \frac{1}{2} G_b)$$

In these formulae, a natural value for the constant of proportionality (the gain factor) is one, which gives you each half the credit (or debit if the excess is negative) for your session performance. However higher values of gain are valid, and are discussed later.

And then, to get our updated grade, we put our newest Session Grade into the right hand side of the decay triangle and get a new weighted average, which is our new current grade.



For Butler, Cross-IMPs and Aggregate Pairs, we have to change the $p - 50$ bit of the formula because we would no longer be dealing in Match Point percentages. This is discussed in the three following sections.

2. Butler pairs

If you think that it might seem a dubious concept or rather controversial to convert IMPs into Match Points, FAQ 18 discusses the question of whether it's a reasonable idea to compare MPs and IMPs, but right here, we need to put that aside and get on with the business!

Fortunately, much of the necessary research had fairly recently been carried out by John Probst in consultation with others, for the production of new Victory Point scales for Swiss matches, particularly for Bridge Club Live. After analysis of shed-loads of data, the key finding is that for any given board, each gain of 1 IMP at Butler Pairs is equivalent to an increase of 5.4% on the Match Point score for the board.

(The comparable Victory Point scales, both for matches of 18 boards, may be found on the Bridge Club Live web site at <http://www.bridgeclublive.com>.)

What we need to do is to take out the $p - 50$ and replace it with 5.4 times the net IMPs for the session or match, for which we use the letter i , and then divide by the number of boards played in the session or match, for which we use the letter x .

Thus

$$SGa = Ga + (5.4i/x + SOpp - \frac{1}{2} Ga - \frac{1}{2} Gb)$$

More recent analysis (please see the Further Reading section (FR 18)) suggests that a gain of 1 IMP at Butler Pairs is equivalent to an increase of 5.10% on the Match Point score for a board.

3. Cross-IMPed pairs

Why the separate chapter? Surely IMPs are IMPs? Well, no, not exactly!

To illustrate why, let's look at a couple of examples.

Say we are playing in a very large field and we make 3NT vulnerable with half of the rest of the field making for +600 and half going one off for -100.

At Butler Pairs, there will be a datum of near-enough +250 and we get +8 for making 3NT.

At Cross-IMPed Pairs, we score a swing of +700 points, which is 12 IMPs against half the field and no swing against the other half of the field, so on Cross-IMPs, we get +6 for making 3NT.

So we get about three quarters of the magnitude of swings for Cross-IMPs compared to Butler IMPs? Well, that was a rather extreme example and here's an extreme example the other way:

Let's say instead that all the rest of the field went off in 3NT while we made it. For Butler, the datum is -100 and we score +12 IMPs. For Cross-IMPs, we score +12 IMPs against every other pair and again, we take +12 IMPs for being the only pair to make 3NT.

So sometimes the swings can work out the same.



OK, so we need a ratio for an average difference in magnitude of the two types of IMPs. Fortunately, this has been thoroughly looked into fairly recently and further discussion can be found in Section 5 at <http://www.bridgeguys.com/pdf/VPScalesJohnProbst04.pdf>

We will use the ratio, known as the Bavin Factor, of 1.2. It's empirical, and the gurus are confident that it's sound, but that didn't stop us from testing it further and analysing around 1000 comparisons from actual results in Bridge Club Live. We got a figure of 1.193, so we were suitably impressed and are very happy with the Bavin Factor of 1.2.

So, for an ordinary Cross-IMPed Pairs movement, we increase the 5.4 ratio by a further factor of 1.2 (which comes to 6.48). So, we replace the "p – 50" bit in the MP Pairs formula with:

$$(6.48i/x)$$

Thus:

$$SGa = Ga + (6.48i/x + SOpp - \frac{1}{2} Ga - \frac{1}{2} Gb)$$

More recent analysis (please see the Further Reading section (FR 18)) suggests that a gain of 1 IMP at Cross-IMP Pairs is equivalent to an increase of 6.12% on the Match Point score for a board.

4. Aggregate pairs

While much research on IMP scoring has already been carried out recently, particularly as a result of many new formats of competition being created for the online club, Bridge Club Live, aggregate-scored pairs lingers on as a bit of a dinosaur. However, it is still popular at quite a few bridge clubs around the country and it's been necessary here to start from first principles.

Full details of the research carried out may be found in the Further Reading section (FR 9).

The replacement for the "p – 50" bit in the MP Pairs formula is:

$$((g-m)/8.2x)$$

where g = the aggregate score the pair achieve and m = the mean score for that direction, and x = the number of boards played in the session.

Thus

$$SGa = Ga + ((g - m) / 8.2x) + SOpp - \frac{1}{2} Ga - \frac{1}{2} Gb$$

So, to summarise the four different types of scoring for Pairs events,

	Score	Average	Contribution to formula
MP Pairs	p%	50%	p – 50
Butler Pairs	i imps	0	5.4 i / x
Cross Imp Pairs	i imps	0	6.48 i / x
Aggregate Pairs	g points	m points	(g – m) / 8.2x

Here, x = number of boards played.

Calculating event grades for teams of four events

There are no plans to include within the NGS results from head-to-head teams-of-four matches, where boards are played at just two tables, as the



scheme would be unable to differentiate the relative performances of the two partnerships within the team.

However, we now include results from many other types of Teams events as it is possible to analyse Teams games as IMPed Pairs. This is because we are then able to reduce significantly the impact of the actions of our team mates from the calculation of our grades.

Such analysis is possible for Multiple Teams-of four events and Swiss Teams events and some Teams league events where Cross-IMP scores can be obtained from the scoring program. These include some Teams-of-eight matches, which are common in County Leagues, and we analyse the Cross-IMP scores as if the event were a Cross-IMPed Pairs event - indeed, for a long time, Butler IMP scores have commonly been calculated for matches between large teams to assess relative performance by the various pairs within a team.

Grades for partnerships

In addition to providing individual grades, the National Grading Scheme provides a number of partnership grades.

The number of your partnerships that are graded as a partnership depends on how often you play with a particular partner. The criterion is that you need to have played 300 boards with a particular partner during the most recent three years. The calculations will thus take account of at least 300 of these boards up to a maximum of 2000 and will use the weighted average of the Session Grade achieved by you and your partner for each event considered.

The formulae used are mostly the same as for an individual's grade, with two differences.

When calculating the partnership's expected performance, the current partnership grade is used rather than the average of the two individual grades.

The initial grade for a partnership, set the first time the partnership plays in a graded session, is not the default individual grade, but is the average of the two players' current individual grades.

Players who have always partnered each other, will have identical grades and an identical partnership grade. Therefore using partnership grades to determine how well you gel with a particular partner will only be possible if you and your partner also play often with several other players, preferably both with some in common, and often enough to iron out random fluctuations.

A special case arises where the two players in a partnership each play more than about 85% of their bridge together, and the remainder of this section is devoted exclusively to these cases.

We start with the formulae for calculating individuals' grades at MP Pairs

$$SGa = Ga + \text{gain} \cdot (p - 50 + SOpp - \frac{1}{2} (Ga + Gb))$$

And equally, for your partner

$$SGb = Gb + \text{gain} \cdot (p - 50 + SOpp - \frac{1}{2} (Ga + Gb))$$

Note that the difference in the session grades for you and your partner is just the same as the difference in your grades before the session. Suppose that sessions with other partners have indicated that you are weaker than your partner "b" (which is never true in practice of course!). When you play with "b"



again, this difference will inevitably be maintained, although not at exactly the same level, since there is nothing in the result of that session to indicate anything to the contrary.

Although this mechanism is logical it can lead to certain anomalies. These are discussed extensively in FAQ 28 in the section “Frequently Asked Questions” and FR 18 in the section “Further Reading” as the difference in the grades between some almost exclusive partnerships has seemed far greater than would be expected.

In particular if both you and your partner each play only a very small proportion of sessions with other partners, then the estimation of the relative strength of the two of you is subject to considerable variability. Whereas NGS grades themselves are not considered mature until a large number of boards have been played, the effect of just a few sessions with different partners can have a long term effect on the relative grades of players who usually partner each other.

For this reason, we believe that the grades of two players who play almost exclusively with each other should only differ to the extent that differences in strength can be reliably estimated from sessions in the recent past of either player. We define a 'regular pair' as two players who have each played over 85% of their sessions (that count toward their grade) with the other. We may adjust the parameter of 85% in the light of further experience. When a regular pair plays together their individual session grades are calculated using a slightly modified formula, viz:

$$SGa = Ga + RPCF \cdot (Gb - Ga) / 2 + gain \cdot (p - 50 + SOpp - \frac{1}{2} (Ga + Gb))$$

And equally, for their partner

$$SGb = Gb + RPCF \cdot (Ga - Gb) / 2 + gain \cdot (p - 50 + SOpp - \frac{1}{2} (Ga + Gb))$$

where RPCF denotes a Regular Pair (or Partnership) Convergence Factor. If RPCF is set to 1 then the Session Grades of 'a' and 'b' will be identical, and this will bring the grades of 'a' and 'b' to the same value, if applied to enough sessions. Any value of RPCF between 0 and 1 will tend to bring the grades of A and B closer together. We have initially chosen to use a value of RPCF slightly below 1 (actually 0.75). This mechanism ensures that when two players play almost exclusively with each other, any difference in grades will tend to reduce over time unless results with other partners consistently reinforce that difference.

Please note that this mechanism was introduced in April 2013 and the effect is gradual, as it won't have its full affect until it has applied to all the sessions used in the calculation of A and B's grades.

Grade initialisation

We need an initial grade for the first time that we play in a graded event for the purpose of other players' grade calculations, including those of partner. For this purpose, we initially adopted the simplistic view of assigning an initial grade of 50% to all players. Once the National Grading Scheme became established and most of the newcomers to the scheme are also newcomers to bridge, the initial grade was gradually reduced to a value of 42% (in October 2012) to reflect this inexperience, and this remains under review. The exception to this is where a player's first event is a Green Pointed event. For such players, we assume that they gained experience in bridge outside

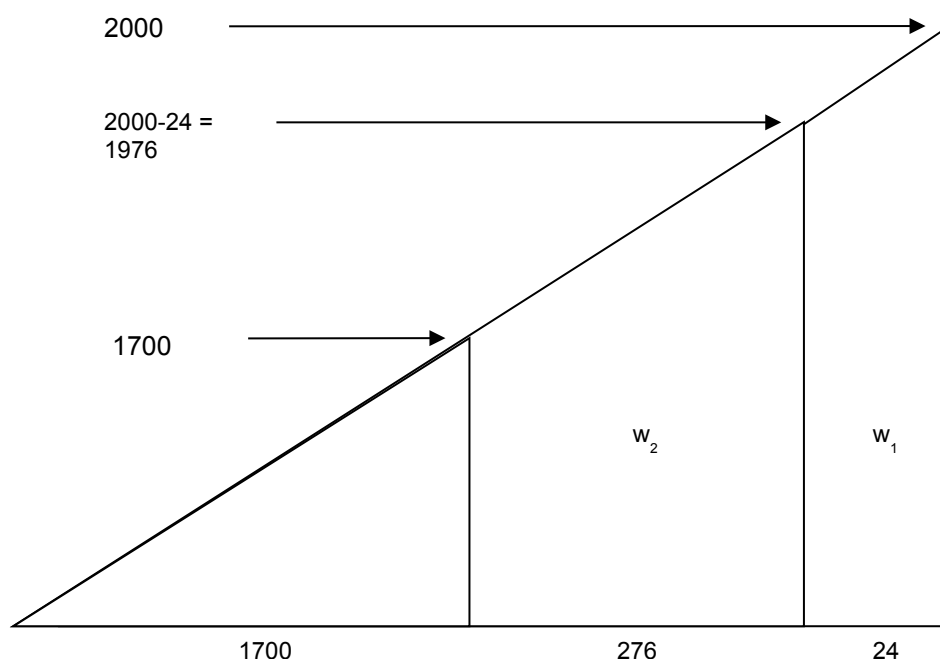


Full Description of the National Grading Scheme

England, and we currently (since October 2012) assign 55% as their initial grade.

If our grade during our first few sessions were to be based entirely on our results for those few sessions, the high volatility of our grade would be undesirable for the purposes of other players' grade calculations, again, particularly those of partner. Therefore, until a player has played his first 300 boards, we have introduced an artificial dampening of volatility to extend a player's history of fewer than 300 boards to exactly 300 boards and saying that he has his initial assigned Session Grade for the boards he hasn't played.

For example, after one session of 24 boards, the decay chart would look like:



The current grade = $(a_1w_1 + a_2w_2) / (w_1 + w_2)$

where a_1 is the Session Grade for the first and only session so far played, of 24 boards,

a_2 is the initially assigned grade,

w_1 is the area of the trapezium on the right, with width = 24 and average height = $(1976 + 2000) / 2 = 1988$

(so that $w_1 = 24 \times 1988$),

and w_2 is the area of the second trapezium from the right, with width = 276 and average height = $(1700 + 1976) / 2 = 1838$

(so that $w_2 = 276 \times 1838$).

The triangle at the bottom left has no assigned Session Grade and so is ignored.

As described elsewhere, a player's grade is not published in the period of the first 300 boards. Once 300 boards have been played, no artificially assigned Session Grade is used in the computation of a player's grade and, depending on the value assigned to the gain factor (see below), the effect of the initial grade on the calculated current grade will be small.



Diffusion

Diffusion is the process by which different clubs and different events achieve different strengths of field. It is the process of players playing in different clubs and at different events, and tending to raise their grade at one venue and lowering their grade at another venue until eventually the differences in the strengths of the fields have reached the right level.

If the members of Bridge Club A are isolated, i.e. they always play within their club and never play elsewhere and never have visitors, their strength of field will always be near enough 50%, yet compared to the rest of the country, the standard of play may be stronger or weaker and we won't really know.

Let's say my wife and I are members of a similar isolated bridge club, "B", in the next village. Suppose "B" is not as strong as "A", and that, because we are keen players, my wife and I join "A" as well. Our grades will tend to go down when we play at "A", and then when we play back in "B", they will tend to go back up. And there will be a tendency for this to keep happening.

Now let's consider what is happening to the rest of the members of these two clubs. As we play in "A" and we lose a bit of grade value, we are spreading that little bit of grade value to the members of "A" and therefore increasing the strength of the field of the rest of that club. Then when we go back to club "B", we tend to gain back the grade value that we lost in club "A" and so we very gradually reduce the strength of the field of the rest of club "B".

So, as we continue to play at both clubs, a differential in the relative strengths of the two clubs appears which will continue until a representative equilibrium is approached.

Most bridge clubs have a large number of members that also play at, or are members of, other clubs and most bridge clubs have a fair number of players who participate in County and EBU events.

The greater the diffusion of players, the sooner realistic relative strengths of field becomes established. For many clubs, diffusion took place very quickly, sometimes within months.

Nevertheless, an empirical measure of the diffusion into a club from players who also play elsewhere has been developed and this gave the NGS a way of estimating how long, if any longer, it will take for each club to have grades which are truly comparable with other clubs in their region. It seems likely that there are a few clubs, probably less than 5%, whose members continue to be over-graded on average. This is slowly being eliminated as new members join these clubs, as they are assumed to have a low grade (as mentioned earlier).

Assisting diffusion

In the early years of the National Grading Scheme, two reasonable methods of assisting diffusion are being deployed.

For the first method, we can change the gain factor in the basic Match Pointed Pairs formula

$$SGa = Ga + \text{gain} * (p - 50 + SOpp - \frac{1}{2} Ga - \frac{1}{2} Gb)$$

from its natural value of 1 to a higher value.

For example using a factor of 2 gives the session grade of both players the full credit (or debit) for their performance in a session.



This factor is called 'gain', by analogy with electronic or digital filters. (We have also called this the 'volatility' factor.) By increasing the gain, we increase the speed at which the system responds to discrepancies in the 'measurement', either in SOpp or in Ga and Gb. So a higher gain improves diffusion and reduces the time for the effects of the initial grade to wear off. It also reduces the time it takes for your grade to respond to a step change in performance, maybe after that course of lessons.

However there is a downside. The current grade is also more affected by random variations that naturally occur from one session to another, and although the linear weighted averaging process reduces these in a complex way, increasing gain increases the volatility of your current grade.

The second method of assisting diffusion is to take EBU and County events, where players from many different clubs come together to play, and to enhance the weightings of the Session Grades achieved by players in these events.

"Type 3" events – Factor of enhanced weighting = 2 (was 3 till April 2013):

Defined as: All events for which Green Points are awarded, but excluding Simultaneous Pairs events.

"Type 2" events – Factor of enhanced weighting = 1 (was 2 till April 2013):

Defined as: All events for which Black Points are awarded on a higher scale than the Club scale.

There are pros and cons relating to the introduction of these enhanced weighting factors and these are discussed in the Further Reading section (FR 6). These factors remain subject to review from time to time.

Limitations to diffusion

Any scheme for producing individual grades across a wide population of bridge players will have some limitations, as mentioned in FAQ 11. For example, the system will not distinguish between the strength of two players who only ever partner each other.

For clubs that are relatively very isolated, the average grades of members of that club would remain for a long time near the original national average of 50%, irrespective of the actual strength of the club's players, as diffusion may be very slow to make much impact. Despite these and other potential anomalies, the National Grading Scheme produces an estimate of individual playing strength, relative to the whole EBU membership, for the majority of players.

These factors remain under review during the early years of running the National Grading Scheme.

Chronology

When data from an event reaches EBU HQ, it is used to update grades. But what happens if an event in which you played a several days before gets scored later? Well, when the data for the earlier event arrives, it gets incorporated for another updating, but as for how, there are two underlying principles:



The two events get swapped round for the decay considerations, so that the pecking order for the decay of event grades is strictly chronological with date of play and not date of scoring.

However, we don't do any recalculations on the more recent event because the effect of changes of grades resulting from the earlier event (that was scored later) has a comparatively small effect on the later event (that was scored earlier).

Similarly, if corrected results for an event are submitted, the erroneous version is first removed and each participant's current grade is recalculated, based on the sessions excluding this one. Then the corrected results are processed using the just calculated current grades. If other sessions for some of the participants have been processed since the original erroneous submission, there will be a small impact on the calculated current grades.

To ensure that NGS processing can be replicated if necessary, each session used by NGS is given a unique NGS sequence number.

Using the national grading scheme for club handicapped events

Many bridge clubs have handicapped pairs events and the National Grading Scheme can provide Club Secretaries the best available ready-made handicaps for players. Each partnership's handicap is the average of the grades of the two players, subtracted from 50%.

A stumbling block will occur if a club member makes their grade "private", but logic would dictate that it would be unlikely that a player who wishes their grades to be kept private would wish to participate in a handicapped event.

The National Grading Scheme thus provides scope for Counties and other organisations also to construct ranked, flighted or handicapped events based on players' current grades, should they feel that any such events may be popular.

Certainly at club level, most players enjoy the boost of having a session when they score better than their average. Handicapped ranking lists enable club players who would otherwise usually be in the bottom part of the field to finish higher up the list on these occasions, and thereby heighten their enjoyment of the game. If your club has a sweepstake, it could be handicapped, or perhaps handicaps might be used for club party or celebratory events.

You cannot handicap Swiss Pairs events unless you were to handicap each individual match, and this is not advised. A better approach, if the object is to award a prize to a non-expert pair, is to stratify the event by having one or more stratifications of pairs whose average grade is below a level or levels of your choosing, and having a prize for the highest placed pair in that or those stratifications.

For handicapped teams events or matches, please see the Further Reading section (FR 7).

Further information on the scheme

The section "Frequently Asked Questions" contains answers with further information on the scheme.



Full Description of the National Grading Scheme

The section “Further Reading” contains further explanation and discussion on various aspects of the scheme



Frequently Asked Questions

At the beginning, we guessed some of the things we would get asked. We have since been able to add some questions that have actually been asked!

Some of the answers provide further information on the scheme.

1. What does my EBU numeric grade mean?

Your numeric grade (e.g. 51.67) is the scheme's estimate of the percentage score that you would achieve on average if partnering another player with the same current grade at Match Pointed Pairs in a field of nationally average strength.

An average strength EBU player will therefore have a strength of about 50.00, but there are no limits to the grade value. The strongest players will have grades of over 60 and the weakest under 40. The numeric grades are divided into grade bands.

2. How are grade bands and numeric bands related?

EBU grades are divided into 13 playing card bands, ranging from 'Ace' at the top to 'Two' at the beginners' end. Each band covers a 2 point range of grade values, with 'Eight' having a range of 49-51. In addition to identify the highest graded players the Ace band (and only this band) is divided into four suits. The full set of bands and their numeric equivalents are shown below.

	grade range
Ace of Spades	67+
Ace of Hearts	65-67
Ace of Diamonds	63-65
Ace of Clubs	61-63
King	59-61
Queen	57-59
Jack	55-57
Ten	53-55
Nine	51-53
Eight	49-51
Seven	47-49
Six	45-47
Five	43-45
Four	41-43
Three	39-41
Two	under 39

3. How are EBU grades calculated?

For each event submitted to the EBU that you play in we calculate a 'par' score for you and your partner for that session, based on your existing grades and that of the other players in the session. If you score above par your grade will go up proportionately, if you score below par your grade will go down.



To find your par score, we take the average of you and your partner's grades, call this GPair. We also find the average grade of all the players against whom you are competing in the event, call this GField.

You and your partner's par score at match pointed pairs is:

$$(50 + \text{GPair} - \text{GField})\%$$

In calculating your revised current grade we use a method which uses your performance over the last 2000 boards you have played, giving more weight to recent sessions. The full description is too long for this FAQ, but is given in the EBU document "EBU - National Grading System - Full description".

4. How many sessions do I have to play to get a grade?

We count the number of boards you play in each graded session. Your grade is provisional and is unpublished until you have played 300 graded boards. This is because it would be heavily weighted towards the default initial grade. After that your grade is called "Evolving" (and denoted by an E after the grade). When you have played 1000 graded boards within the last three years, your grade is called "Mature" (and denoted by an M after the grade) and from then only the sessions making up your most recent 1000 boards will be used in the calculation of your grade.

If you subsequently take a break from playing so that you have played less than 300 boards in the last three years your grade is called "Out-of-Date".

5. Will my grade change by the same amount as that of my partner after a session together?

When you partner a particular player, your grade will go up or down by very similar amounts following the boards that you play together.

The most likely reason why the amounts may be very slightly different is due to the method of decaying of the effect of your previous results if you have played with different partners over your previous 2000 boards.

If you and your partner each play over 85% of your bridge with each other, in April 2013, a mechanism was introduced to slowly converge your grades.

6. How does the effect of a board or a session that I play today diminish over a period of time?

The speed at which the effect of a past session diminishes will depend on how often and how many boards that you play every week or every month. The more you play, the faster it diminishes.

If you play two evenings each week at your local club and play a 25 board movement each time, then the effect of one session will halve in 20 weeks and disappear in 40 weeks. If you average just one such session per week, decay will take twice as long.



7. How can I work out, as I start to play an evening duplicate at my club, what % Match Points score I need to achieve in order to maintain my grade at its present level?

This is yours and your partner's 'par' score as described in FAQ 3 - "How are EBU grades calculated?" This will depend on the grades of yourself, your partner and the opponents.

With a strong partner and weak opponents, you may need to score 60 or 65% on average per board to maintain your grade, while with a weak partner and strong opponents, you may be able to maintain your current grade just by scoring as little as 35 or 40% on average per board.

In short, the system is designed so that it doesn't matter who the other three players are at the table with you, nor where you are playing, because the factors of their strengths are taken into account.

8. Is there not an advantage for regular partnerships?

Yes, there is, and, unfortunately, there is no good way of taking this into account. As a generalisation, there is a slight advantage enjoyed by a regular partnership over a new partnership, but any adjustment for this would need to be empirical. It is currently thought that about 2% could be added to the grade of a player who plays mostly with pick-up partners for comparison with a player who plays mostly with a small number of regular partners, but there again, the extent of partnership cohesion or gelling differs between different regular partnerships.

Because of the disadvantage for unfamiliar partnerships, and to avoid discouragement of players to act as "hosts" or "stand-by players" at clubs, it will be made possible for a player assuming that role to indicate before the start of play that the pair wishes not to be graded for that particular session. This facility is limited to one pair per event.

Following analysis of data for the 2013 period, an article written by Michael Clark and which appeared in English Bridge, August 2014, included the following:

"Myth 2: You're disadvantaged playing with an irregular partner

This is another common one and it sticks around because there is some truth to it. If you play only with regular partners you will clearly have a higher grade than some - one of the same ability who plays only with pick-up partners. But *how much* does this matter? We estimate that the difference is about 2% between the two ends of the spectrum, so that means that if you suddenly play with a new partner, having only played with regular partners previously, and your new partner is in the same position, you'll need to get about 4% more than usual to break even.

But this is extreme. In reality you'll play with a range of partners, some more regular than others, and your partner-forming habits will already be built into your NGS grade. When you play with the newer partners, you'll be a tiny bit disadvantaged, but when you go back to playing with the regular partners you'll be correspondingly rewarded.

The facts: In 2013 there were 80,000 partnerships who had played fewer than 50 boards together. Almost exactly 50% of those partnerships improved their grade."



Another known factor that cannot be taken into account by the scheme is if a player plays when unwell or is unable to concentrate as well as normal on the game, leading to a poorer result than might otherwise be expected.

9. Can I opt out of the scheme?

You cannot entirely opt out of the scheme, as your grade is needed in the calculation of other members' grades. You can change the way information on your grade is presented. If you opt to make your grade "Private", then:

Your current grade will be indicated as "Private" to your club.

Your grade will not appear in the published lists on this web site.

Your club would be unable to use your grade for any handicapped events it may hold.

10. Can a "host" player have an event excluded from the scheme?

A "host" player together with his/her partner may opt to be excluded from the grade calculations of any particular event. This is limited to one player or pair per event, is at the discretion of the event organiser, and the pair must express their intention to opt out before the start of the event.

Bear in mind that players new to the grading scheme are assumed to have a below average grade, and the grade calculation for this event for them and their partner will take account of this.

11. Why do grades seem to be so sensitive, and why is so much weighting given to the session just played?

This is a relative matter. Because the figures are recalculated every time that you play, you can track small changes in your figure after every session is graded. It is expected that for most players, your grade will tend to vary within one or two bands (a standard deviation of about 2% in grade value) in the long term.

The weighting is such that each board in the session that you have just played will account for almost one fifth of one percent of your grade value, so a top or a bottom will increase or decrease your figure by approximately 0.05%. This may seem significant but it really isn't. Players should expect their grade regularly to go up or down by 2 or 3%. This is not a great movement given the wide spread of grade values that there is across the whole field of players.

Using more boards to calculate grades would reduce this volatility (or reducing the gain factor in the formulae), but at the expense of a lower responsiveness to any difference between your grade and your current form.

12. Looking at my sessions on the website, why does my grade sometimes seem go down after a good result?

This can happen if you were playing with a strong partner or against weaker opponents or in a relatively weak field.

However, there is a further possibility.



Your sessions are displayed in order of date played, but the NGS grade is recalculated when older sessions are submitted (or corrected) late. So occasionally your grades are not displayed in the order they were calculated, and this can give rise to the apparently anomalous effect.

Allowing for this, your grade will normally go up if you score above par and down if you score below, but your grade can also go up or down, even when you score on par, because older results are being given less weight.

13. I was partnered with a “King” graded player last night, but he/she didn’t seem to play at a level that I would have expected of a “King”.

It could be that your partner was having an off-day, or that you did not gel as a partnership.

14. Why should my grade suffer when partner plays like an idiot?

Bridge is a partnership game. While one or two online bridge providers have attempted to provide empirical methods to assign responsibility within a partnership by weighting differently between the players in a partnership depending on who is declarer, it is our view that such methods are fraught with difficulties. We believe that in the long run, you and your partner will each perform roughly in line with the expectations that are your current grade. And while you may remember that cold game on Board 15 that partner threw on the floor last night, you may have completely forgotten her brilliant play on Board 8 that got your side a complete top.

Note that if partner always plays like an idiot, this will be reflected in his grade, and so the grading system will have a lower expectation of your partnership. Also remember that one attribute of a good bridge player is to bring out the best in partner. Partner is less likely to behave like an idiot if you don’t get annoyed with him.

15. Do I need to play with different partners to have an accurate grade?

Imagine that both you and your partner have never played except with each other. You would have identical current and session grades, even though one of you may be better than the other. If at least one of you plays with other partners, this will gradually allow a differential to develop between your grade and your regular partner’s grade. The speed at which this happens depends on the degree of mixing of partnerships.

It’s also possible to get misleading differences in grades in an example such as three players, A, B and C, where A partners B or C exclusively and both B and C partner only A. We know that this and similar situations do occur, although relatively infrequently across all EBU members. It is recognised that some individuals’ grades will be less accurate as a result of this.

It is intended that these things be monitored and assessed in the initial months and years of the National Grading Scheme. That is one reason why the National Grading Scheme also produces partnership grades, which are less susceptible to this sort of phenomenon.



16. Why does my grade seem to rise in small amounts and fall in large amounts?

This is more likely to be a perception rather than reality. However, a possible reason may be that you were unwell or particularly distracted during a particular session.

17. How often will grades be updated and when are corrections possible?

We currently update grades during the night following submission of scoring data to EBU HQ. This should increase the chance that your sessions will be processed in the order they were played.

However, the system will take account of corrections submitted at any subsequent time, and will recalculate the affected session and the overall grades, as part of the following day's NGS processing.

18. Why hasn't one of my recent sessions shown up in my NGS data?

This could be due to one of several reasons.

1. The session has not yet been uploaded to the EBU website.
2. The event is not one of the types of event currently graded.
3. A very few sessions have data errors, which causes the grading system to reject the session, or to exclude certain pairs when we cannot be certain of their partner's identity.
4. Your partner is unknown to EBU. Only pairs where both players can be identified are included in the NGS processing of a session.
5. There may be an extreme situation in which the session had too few pairs who are EBU members. We can cope with a few players of an unknown strength, but the NGS needs over 2/3 of the pairs to be within the system to process the session reliably.

19. What about Simultaneous Pairs and other multi-section events?

For Simultaneous Pairs events that are scored both at the local club and nationwide, it is reasonably obvious that we should take one scoring rather than both or else we would get a double weighting for the event. So, which scoring should we take?

Local factors at the club can have a significant effect on the National scoring, especially so if a two-winner Mitchell movement is used. Some clubs can have stronger fields sitting North-South than East-West and some the other way round. This can happen without intention but by default, and for different reasons. Because we use a Strength Of Opponents Factor for duplicate pairs movements, the club score rather than the nationwide score provides the appropriate scoring for grading.

Similarly, for multi-section events played at multiple venues, we take the scores for each section separately rather than the combined section scores.



For multi-section events at a single venue, such as EBU's championship pairs, we take the overall score for the event. However qualifying stages and finals are treated separately.

20. How are grades standardised across the country?

This is achieved by the process of diffusion, by players who play at different clubs or in events with a mix of players from different clubs. Please see the section on diffusion in the Full Guide.

21. Should you really be trying to compare IMPs with MPs? In any case, I play better at one than the other.

One argument runs that we should have two different grades for each player, one for MP play and one for IMP play. That argument is outweighed by the view that we really seek to have just one form of grading for the game of bridge and that it would be somewhat cumbersome to have different grades for different types of scoring. Maybe you play more of one type of scoring than another. If so, your grade then reflects your overall standard based on the proportions of different types of game that you have played recently. There are very few players who are considered to be significantly better players at one form of scoring than another.

How confident are we about converting IMP results into equivalent MP results? We start easily enough with a zero IMP swing being equivalent to a 50% MP score on a board and then we need a scale of how many %MPs to 1 IMP. It works out roughly that a +1 IMP (Butler) swing is equivalent to a score in the order of 55.4% MP on a board. But there are IMPs and there are IMPs, and Butler and Cross-IMPing and Teams-of-Four IMPing all have different characteristics that need separate consideration. Interestingly, much of the Maths groundwork on all this had already been carried out in the formation of Victory Point scales that are used for matches and teams events.

We are not the first to produce such conversions. OK Bridge, an online bridge provider based in the USA used similar conversions for their grading scheme for some years before us.

22. I perform better in a strong field / in a weak field / with a stronger partner / with a weaker partner. Can I beat the system?

Such views are expressed with some frequency but to a large extent the perception does not match reality. Where there is an occasional element of reality, then, yes, you may find that you can optimise your grade by playing more with a particular partner or in a particular standard of field, but the advantage is unlikely to be any more than very slight.

You can beat the system, by playing with a partner who you are sure is under-graded, or by playing in a session against players who are generally over-graded. Such partners and opponents will exist (see the FAQs on the accuracy of your grade and on grade standardisation) but the fact of your playing with them will help to eliminate these errors, so we positively encourage this!!



Of course, the best way to improve your grade is to improve your standard of play, for example by taking lessons, reading books, and playing with strong players who'll coach you.

23. Do other countries have similar grading schemes?

There are player ranking systems in many other countries. These vary widely; from systems very like our Master Points system, some with automatic reductions of points gained in earlier years, to systems that use percentages like our grading system. We believe our system is unique, in that it is not biased towards players who play more frequently, it takes account of all competitive games played at all levels of the EBU in the same manner and so includes all EBU members, and it is focused on current performance

24. What about when I have a sit-out?

Only boards that you have played will count towards your session grade. In the same way, sit-out boards do not count towards your results for a session as you are scored on the boards that you do play. If you played a 27 board movement at your club last night but sat out for three boards, the system will count you to have played 24 boards.

25. What happens when I am given an “average” on a board?

This is different to a sit-out because the board is one that you were scheduled to play and you did receive a score for the board, even if it is an artificial adjusted score if you were unable to get a playing result. In this case, the board counts towards your total for the event, both for scoring the event and for grading calculations.

26. Why do many of my recorded sessions indicate that I played 24 boards when I know I played more or fewer boards?

This may happen very occasionally and is because the scoring software that was used at the events currently does not report the number of boards played. Thus a figure of 24 boards is used as default in these cases.

27. When will Teams results be included in the NGS?

Many Teams results are now included.

For Multiple Teams of Four, Swiss Teams, or League events where each board is played by several pairs, we calculate the performance of each participating pair (using Cross-IMP scores for each pair). So if the scoring software for such an event produces a set of Cross-IMP scores for each pair and identifies which pair were NS and EW for each board, the grading system uses the results from these events.

There are no plans to include within the NGS results from head-to-head teams-of-four matches, where boards are played at just two tables, as the



scheme would be unable to differentiate the relative performances of the two partnerships within the team.

28. I mostly play with my regular partner, Ann, and expected to have almost the same grade. Why is my grade so much less than hers? Surely there's a mistake?

This is without doubt the most significant issue raised. In one or two extreme cases, differences between the grades of two frequent and reasonably strong partners have reached up to 20% such that one partner is an Ace Of Spades while the other partner is an Eight, leading this to be dubbed as the “Aces And Eights Issue”.

A reply follows here, and a couple of case histories are discussed in detail in the Further Reading section at FR 13.

This situation occurs to some extent to some hundreds of regular partnerships, and is important to them. So a longer than usual explanation is appropriate.

No, there's no mistake in the grading system. It tries its best to determine your individual grades based on the sessions of each of you with other players and can make an estimate with only a few such sessions. There are several possible reasons why your grade has come out lower.

1. Your lower grade reflects a real difference between your and Ann's performance in the last couple of years, either due to your poorer than expected performance with other partners or Ann's better than expected results with her other partners.

2. You have been unlucky, or Ann lucky, when playing with other partners. Differences of around 2 points (one grade band) are almost certainly due to chance, and up to 4 points are not uncommon.

3a. Your other partners are weak players, and NGS has not recognised their weakness. (But note: your grade won't go down playing with weak players who are correctly graded.) This can happen, but we try to minimise it by assigning a below average initial grade to new players.

3b. Ann's other partners are stronger players, and NGS has not recognised their strength. This can happen, especially if they play with very few partners, though it is less common, as many stronger players do play with a number of partners. We can't mitigate this, but the effect would go away if Ann could persuade them to partner you instead a couple of times.

4. You just don't "gel" with your occasional partners, and so your results don't reflect your normal strength. We know this is often a real effect but (sadly) can't quantify it. If this is a result of your acting as the "host" for a session, you will be able to ask your club administrator, at the start of the session, to mark you in this session as non-rateable (only one pair per session). Sorry, we can't do this retrospectively. Also, persuade Ann to do her share of playing with new members!

We can't identify which of the above reason applies to you. It may be a mixture of several of the above. However we have analysed a number of cases as part of testing the system. In the majority of cases the reasons are (1) and (2). If the reasons are (3) or (4) in your case, please accept that you are in the minority, no bridge individual grading system can be accurate for 100% of players. You can instead focus on your partnership grade, as this is based solely on your and Ann's grades together. We'll be looking at ways of making partnership grades of regular partnerships more publicly visible in the future.

On a similar subject, we have been asked if we can introduce “some factor that slowly converges pairs within a partnership”.



There are arguments against this.

1. It penalises you whenever you play with a weaker partner, which would upset a different set of players.
2. For a frequent partnership, when playing with occasional other partners, this is the only time for the scheme to try to achieve a realistic differential between the grades of the frequent partnership. For all we know, the scheme has found and continues to find the right differential.

Nevertheless, the NGS started, in April 2013, to bring closer together the grades of about 3000 pairs who were partnering each other over 85% of the time, but who had occasionally partnered someone else. There were about another 1000 pairs who had only ever partnered each other in graded games, but they were unaffected.

29. Non-graded sessions

We were asked:

"I have an entry against my 2012 sessions on the EBU web site for 6th February, which gives me 60 master points but has what I assume is an artificial score of 0% listed. How do you filter this type of scoring blip out of the scores that make up individual NGS rankings?"

Various sessions are ignored by NGS: see FAQ18.

30. Teaching sessions

We were asked:

"Should sessions be included in the NGS calculations if those sessions are 'novice' or 'teaching' sessions? By this I mean sessions where a teacher is on hand to give advice about bidding or play. Many such sessions seem to have been included, which must be wrong."

Teaching sessions are now excluded. Initially clubs had no method of identifying such sessions.

31. Student and "Pro-Am" sessions

We were asked:

"I currently play at our local club on the 2 strongest nights - Monday and Thursdays.

Saturdays are called "gentle duplicate" and tend to have the less able players.

Wednesday is "improvers" duplicate (previously students duplicate) and is aimed at the players who have been to lessons, but never played in a duplicate.

I am planning to start playing regularly with a friend, who has never played duplicate before, on Wednesday nights. I predict that our results, certainly initially, will be pretty random. I know several other more experienced members also play with students on Wednesdays to help them.

I am not clear how the ranking system can do anything but penalise those of us who do this, which is in the aim of supporting less experienced players being introduced to the game. You state that one pair per session can opt out.

However, wouldn't it be fairer for the whole student session of the club to not count?"

'Teaching' sessions are ignored by NGS if they are declared as such by the club, but sessions with beginners are included if not 'supervised'.



The NGS does take account of partnerships being of different strength. This should not be confused with the fact that it can't take account of the difference in performance between a regular and a pick-up partnership (see FAQ 8). It's a good idea to have the whole student event count to mature the grades of the students, and there should be no disadvantage for stronger players to play in such a session. The questioner opines that his results with his less experienced friend will be "random". It is unlikely that these results will be any more "random" than with any other partner as there is no evidence that their results cannot be predicted as per FR.5 (Further Reading Subject No 5)

32. Novice sessions

We were asked:

"The effect you describe of "isolated" clubs maintaining a SOpp of 50% whatever their objective standard is potentially a problem within individual clubs where there are sessions restricted to particular classes of member, such as novices, and little mixing with the rest of the club (not necessarily club policy, merely custom). If, for example, a single member plays both in the novice sessions and the expert sessions, he will become the partner to avoid in the expert sessions. His rating will top up in novice sessions and drop in the expert sessions for some considerable time while rating diffuses away from the novices to the experts through him. The outcome for him is neutral, but for those partnering him in the expert sessions, it is a sure recipe for systematic grade deflation."

This is true but necessary for diffusion. In practice, the job of partnering the mixer is usually rather more shared out. Those 'experts' who become under-graded will be highly sought after partners for the other experts, so things should even out.

Most situations of this kind will have taken place a long time ago, as the initialisation of the scheme was April 2010. We have found generally that where clubs have different sessions of differing abilities, diffusion has acted quickly in providing reasonable respective Strengths of Fields.

33. How is diffusion progressing?

Most diffusion took place very quickly, in some cases within a month of the start of the scheme, but we retain some settings to encourage continuing diffusion to come.

We received one missive saying: *"This is a disappointment. It is obvious that the club cannot use this for handicapping as the ability rankings are wrong in the majority of cases, sometimes wildly so. e.g. Bob is 54% and Jean 52.9%. Bob is a beginner (club master)."*

The NGS is complex in its means of calculation but perhaps what it does not recognise is that club sessions are of a different standard. We never get our Grand Masters playing on a Monday or in the No Fear etc. The weaker players get a higher ranking as they never meet stronger players etc."

Upon investigation of this case, we discovered that diffusion had already worked here. Jean was over 59% in June 2011 and again so at the time of writing in April 2012! She plays in high quality fields. Bob had not played since June 2011 and had an exclusive partnership with Peter who has exactly the same grade. They played in below average fields throughout.

34. Stratification and NGS

We were asked:

"Having read the National Grading Scheme paper on the web site I see that there is no mention of stratification. When it was introduced I understood that it would be possible to set the Strata by National Grading Scheme Grade or by Master Point Rank. Is this still the intention?"



The NGS Working Group has not considered this as a high priority. The intention still exists, though we will be guided by the views of the EBU Tournament Committee. While the NGS creates this opportunity for stratification, some may argue that stratification for the awards of Master Points should be based on Master Points already held, and that the NGS is better suited as a basis for handicapping. No doubt, this and related subjects, possibly even the EBU Ranked Masters Pairs, are expected to be debated in the times to come, but these are matters more appropriate for consideration by the EBU Tournament Committee rather than the NGS Working Group.

Meanwhile, we have already heard of events being held by clubs in which prizes have been awarded to pairs using stratification by NGS grades.

35. Converting NGS to Handicaps

We were asked:

"We use Scorebridge to score our handicap events and Bridgeweb to display the results. You address how to manually convert the grade to a handicap, but has an interface been written to allow automatic conversion? If not, there would appear to be a significant manual overhead to using the NGS."

That is right – there is currently no automatic conversion. Maybe authors of bridge scoring programs might consider this in due course.

36. How can we set up handicapped events if we have club members opting for privacy?

Club administrators cannot see the grades of members opting to have their grades private, and this will inhibit or deter clubs from holding handicapped events based on the NGS. It is hoped that this fact will persuade many members to participate in the NGS, rather than remain private or even anonymous. (Tournament Organisers would otherwise be best advised to estimate a grade that is private and to try to err on the high side.)

37. "The NGS has meant that I won't play with weak partners"

Following analysis of data for the 2013 period, an article written by Michael Clark and which appeared in English Bridge, August 2014, included the following:

Myth 1: You're disadvantaged when you play with a weak partner

We hear this one quite a lot. Someone will claim that they no longer want to play with poor Mr P, who hasn't scored above 50% since the late sixties, because it will damage their NGS grading. However, they're worrying about nothing. If both players are correctly graded, then it doesn't matter what the individual strengths are – the NGS takes this into consideration.

The maths is quite simple: if Mr P is graded 30, you're graded 50, and you're playing in a field of average strength (50), then your expected score will be just 40%. If you do better than this, your grade will go up; if you do worse your grade will go down.

The NGS should actually make you more keen to play with Mr P. In the old days you had nothing tangible to gain from playing with him – you were unlikely to win any master points or beer vouchers. With the NGS, though, you have a real goal to aim for and can take some pride in eking out a well-earned 42%, to improve both your grades.



The facts: In 2013 there were 1450 incidences of a 60+ player partnering someone with a grade below 45. In 49.7% of those sessions their grade went up

38. Are there plans to export the NGS? Scotland, Wales, further...

Well, that's up to them! However, there is no intention to include online bridge within the scheme.

39. Can we improve the display of the NGS on the web site?

Yes. It is intended to continue to increase the amount of information available

In particular, we were asked:

"Will you publish a table of SOpp values for individual clubs to assist the "diffusion" process you describe, which will otherwise take many years?"

The SOpp (Strength of Opponents) changes from one session to another. It is possible to see the SOpp for each session in which you played via your Member's Area on the EBU web site.

The Strength of Field for each session is available to Club Administrators.

We will consider whether Strength of Field information can reasonably be increased. Some clubs take information from their Club Administrator's page to advertise a recent average of their Strengths of Field for each of their weekly events.

40. How often does the NGS re-calculate and when?

We process results daily, starting at 4.30am.

41. Are clubs generally processing their P2P submissions promptly?

Many clubs now process P2P submissions straight after play. The EBU can do nothing but encourage clubs to submit immediately after the game is scored, but we cannot make people do that. We have been encouraging, and will continue to encourage clubs to submit quickly.

42. Have you had the "gain" turned up well above 1? Has that made it more likely that there is under or over shoot?

"Gain" is the prescribed variable that we use to control volatility of grades. It is described in the Full Guide. Gain is currently set to 2 (as it has been since the scheme was initialised), which helps both diffusion and the rapid attainment of an approximately correct grade for a new member. We have found that it does cause a minor and temporary over-shoot for players who have played between around 250 and 350 boards.

43. Average grade of partnership and partnership grade

We were asked:

"I have a grade of 62.1 and a partnership grade of 59.0 with my usual partner, yet his grade is 54.9 making an average grade 58.5"

Differences can arise between the partnership grade and the average of the grades of the two partners. In routine circumstances this may not give a measure of partnership cohesion. Such differences naturally arise through the calculation of individual and partnership grades.



Here's a short story to illustrate this point.

Twins Amy and Annie play together and each sometimes partners Bill. All results are 50% except that Amy and Annie gel and score 54%. For grading we have data from three partnerships and three grades to be determined. There's a unique answer, the girls' grades are 54 and Bill's is 46. The partnership grades are equal to the average of the partners' grades, so NGS detects no gelling, and bad luck on Bill.

Now this doesn't change if it is revealed that it is sometimes Bill's brother Ben who has partnered the girls. Both Bill and Ben have grades of 46.

But now Bill and Ben start to partner each other (to demonstrate that they aren't so bad) and score 50%. This is 4% above their current average grade and it seems that Bill and Ben gel, whereas in fact it's the girls who gel.

Now we have data from 6 partnerships and 4 grades to be determined. With more data than unknowns, the system is over-determined, and the steady state solution for NGS will depend how often the different partnerships play. In the case where mixed pairs play as frequently as same sex pairs, the grades will settle at 53 for the girls and 49 for the boys, and of course partnership grades of 50 for all except for Amy/Annie with 54. Now it seems that same sex pairs gel by +1 and mixed pairs anti-gel by 1.

We think that in practice, if NGS says you gel better with one partner rather than another, it is likely to contain some degree of truth, but as the above example shows, the truth will be cloaked in a lot of system uncertainty.

44. Initialisation

We were asked:

"You say: "Until you have played 300 boards in graded sessions, your computed grade is still significantly affected by the initialisation process". However, such a grade is still used in evaluating partners. In the initialisation process, the strength of a player can be significantly over- or under-estimated and this can have an unwarranted influence on the grading of other players, particularly those who don't record a large number of P2P sessions. Why not leave grades derived from fewer than 300 boards out of the process altogether?"

It's not right to use a session's data to update one player's grade but not their partner's grade. We need to initialise the new player's grade as soon as we can, so there is no reason to exclude any useable data from the calculations.

Although there are a very few chess grading systems that take some account of the assumed accuracy of you and your opponents' grades when deciding how much to update your grade, it is normal in grading systems to treat all players equally. The extreme of ignoring players' results with partners who have played few graded sessions wouldn't be OK, as there is no magic point at which grades become accurate. It is certainly right to say that a scheme that fully took account of the assumed accuracy of each grade would be much more complicated to implement and for players to understand.

Our principle is to give everyone a session grade for each session that they play, as far as possible."

45. My partner and I used to play occasionally with other people but for the last 2000+ boards we have only played with each other. Why are our NGS grades different?

The short answer is that the NGS guide tells the truth, but perhaps not the whole truth, which is: "Your grade is based on your results over the last 2000 boards **PLUS the NGS grades (relative to your grade) of your opponents and partners over that period.**"

If this weren't true, there would be no way for what we call diffusion to work and to identify groups of players (eg the members of a weak club) whose average grade was below 50.



Full Description of the National Grading Scheme

However it does mean that if the NGS thought you were weaker than your regular partner 2000 boards ago, and you and he/she play together for the most recent 2000 boards, then NGS will have no evidence to change its opinion, and you'll stay with the lower of the two grades.

There is just nothing in NGS which says:

"If a strong and weak player play regularly together, it should be assumed gradually that they are of equal playing strength."

Some people say that there should be such a rule, because maybe NGS's earlier estimates were wrong. Of course this might not go down well with those players who were really stronger than their partner. It might also not be fair to any occasional partners of the weaker player, as NGS will assume they are playing with a rather stronger player.

In summary, if you and your partner only partner each other over a long period, no grading system can deliver an accurate estimate of your individual current performance.

FAQ 28 refers to similar matters.



Further Reading

Welcome to the murky depths of the “Everything you could possibly want to know about the National Grading Scheme” pages! There follows a miscellaneous collection of explanations and discussions on some of the detail of the scheme.

The format is that of Frequently Asked Questions.

1. Is the average grade exactly 50%?

Our initial answer to this question was:

“The average grade will be close to 50% but it can’t remain at exactly 50% or remain absolutely constant because of the way the calculations work. The mere fact that some players play more often than others is enough to knock the average about in either direction.

It is desirable, though, for long term comparisons that the average grade remains close to 50% and in the event of a significant deviation, a recalibration will take place such that the current grade of every member will be increased or decreased by the same amount so bring the average back to 50%.

For each member this will be achieved by applying constant reductions or increases to each of the Session Grades that make up the player’s overall Current Grade.”

However, as at April 2013, the average grade (of all those who have played over 300 boards) is just below 49%, but no recalibration has been made. This will be kept under review, but the average has remained as it is for some while. The average Strength of Field is very close to 50% and this suggests that stronger players play more frequently than less experienced players. It may be that an equilibrium has been reached and any recalibration will eventually lead us back to this equilibrium.

2. What is the difference between “straight line” decay and “exponential” decay, and why have we opted for “straight line” decay?

The most common method of decay of the effect of old results, as used by many other countries and also by Bridge Club Live, is what we call “exponential” (or “radioactive” because of the similar pattern to radioactive decay of materials). This may sound complicated but, strangely, it is relatively very simple. The principle is that after an event, your new grade is comprised of a portion of around 5% of how well you did in the most recent event, and about 95% of the grade that you had before the start of the event. The proportions vary from scheme to scheme and may vary within a scheme for various reasons, but the main point is that the effect of all the previous events goes down by the same proportion, somewhere around 5%, after every new event.

When you keep taking about 5% off the remaining effect of an old event, you’ll find that the effect is halved (from what it was after you just played it) after you have played about another 13 events. And the effect will keep halving after every further 13 events that you play. And after two years of playing once a week, the effect will be very, very tiny – about 200 times tinier



“Straight line” decay sounds simpler but is actually more complicated to program. As described in the main text, the effect of an old event goes down constantly by equal amounts every time you play a new board, and when you have played 2000 new boards, around 80 sessions of bridge, the effect will disappear entirely. If you play twice a week, the effect of an event that you played over 40 weeks ago will have disappeared entirely.

So which method is better? There are arguments for both and these could fill volumes of debate, but in the end, there are no “rights or wrongs”.

We have chosen “straight line” decay because there was a strong desire that the effect of old events should be seen to disappear explicitly, and this was considered an advantage sufficient to outweigh any disadvantages.

Finally, on a rather technical note, while “straight line” decay “does what it says on the tin” with regard to the weight of Session Grades, each result affects both partners’ current grade which then affects their Session Grade for all subsequent events. This changes the straight line decay for a particular session slightly. The effect depends on the partnership patterns of each player, but in all cases the effects of old results are negligible.

3. Can a single board have a greater impact on my grade if some other scoring than Match Points is being used?

Yes! OK, then: **Why?**

One of the features of Match Point scoring is that every board is as “important” as every other board. You can score 100% on a board for being the only pair to make an overtrick in 1NT or for being the only pair to make a Grand Slam when everyone else has gone off.

When you are playing IMP pairs (let’s say it’s Cross-IMPs for example) and you score +20 IMPs for (the very extreme example of) making that vulnerable Grand Slam when everybody else went off, you get the equivalent of 179.60% on that board, compared with 100% at Match Points, which makes it worth more than two and a half outright Tops. (The reason that it is 179.60% is that we have a conversion rate of 6.48% of a Top per IMP so we add $20 \times 6.48\%$ to 50% and get 179.60%.)

Now, you might say “Well, that’s not fair, let’s put a ceiling on that so you can’t get more than the equivalent of 100%”, but you would need to consider the other board where an opening 1NT is passed out at every table and everyone else makes exactly while you are the only pair to make an overtrick. Here, you get 1 IMP and the equivalent MP score of 56.48%, and the scope for getting anywhere near an 8 IMP swing on this board to break the 100% barrier seems very remote.

So we justify not capping the number of IMPs that can be won on a board by saying that we need the swingy hands to have a greater impact than a board at Match Points in order to balance out (or compensate against) the flattish hands where there are few opportunities for significant swings, and we need to do that because we need to be able to compare typical sessions of one type of scoring with another.

Interestingly, at Aggregate scoring, which is far more equivalent to the original Rubber version of the game, the flat boards are even flatter and the swingy boards even swingier. The overtrick in the 1NT hand would not be worth enough to feed the chickens while the extreme example of the Grand Slam



swing (with a 100% score in MPs being equivalent to 409 aggregate points above the mean on a single board) would be worth about 333% in MP equivalent, or about five and a half outright Tops.

4. Would it be reasonable to put a limit on the influence of an exceptionally good or bad session?

Because of the random nature of the sequence in which we get good scores and bad scores, there are times when we feel we can go for a whole session without being able to get anything right, and others when we feel as if we can do nothing wrong. There is no mathematical justification for capping the influence of particularly good or bad sessions and every chance that if we were to attempt to do so, we may not get the balance correct, and it would certainly be unfair for any player in the process of improving as we would expect such a player to get more extremely good scores than extremely bad scores. So the answer is “No”!

5. Can I use the information from the National Grading Scheme to work out the probability of how I will score this evening at my local club?

Yes! First of all, you can work out your expected average score by subtracting the strength-of-field of your club from the average of your and your partner's current grades, and adding 50%. So if your grade is 50% and your partner's grade is 46% and you are playing in a club where the strength of the field is 49%, it works out that your expected average score will be $48\% - 49\% + 50\%$, which is 49%.

Analysis by Barrie Partridge at Sheffield Bridge Club some years ago showed that the spread of results for an evening's duplicate, achieved by a partnership in relationship to their expected average score, was that of a Normal Distribution around that average, and that this Normal Distribution was otherwise defined by the standard deviation of approximately 6%.

The great thing with Normal Distributions is that once you've got one, you don't need to bother with maths, because you can look up all you need from mathematical tables (are you old enough to remember Log books at school?) to arrive at some interesting figures.

For example in the case where our expected average score is 49%, we can say:

We have about a one-in-forty chance of scoring over 61%,

We have about a one-in-six chance of scoring over 55%

We have about a one-in-three chance of scoring between 49% and 55%

We have about a one-in-three chance of scoring between 43% and 49%

We have about a one-in-six chance of scoring less than 43%

We have about a one-in-forty chance of scoring less than 37%

The original analysis was carried out for sessions where there were around 10 to 18 tables. For clubs with fewer tables, there is a tendency for results to be more extreme and therefore a greater probability of achieving results that are



further away, in both good and bad directions, from your expected average score.

Accordingly, a more recent study in 2014 by Mike Christie using NGS data found a standard deviation of 6.34% (please see the Further Reading section (FR 18)), though this is close enough to that of the original study that the round figure forecasts from the Sheffield analysis remain valid.

6. What are the arguments for and against introducing additional weighting for certain events?

For the start of the scheme and possibly indefinitely, though subject to review, EBU and County and similar events will have an increased weighting. Note that there is no upper limit to the grade you can attain, even if you never play in such events, provided you score highly enough in club events. The additional weighting just increases the speed at which your grade is changed by the results in these cross club events.

The primary reason for the extra weighting was to aid diffusion (and so standardisation of grades) in the early years of the scheme.

However, other arguments exist.

One view is that when we play bridge we are always trying to play at our best, even when we are not playing at more prestigious events. The counter-view is that some players will happily attend a mid-week club evening session when tired after a day's work, but would be careful to be fresh before a weekend national event.

A disadvantage of event weighting is that the complications to the maths of introducing another scheme of weighting (on top of that of decay) make it more difficult for the ordinary player to understand how weighting and decay work or why their grade is doing what it is doing.

7. How might we handicap team events using the National Grading Scheme data?

Following from the section on Cross-IMPed Pairs, we expect a pair playing against a pair with average grade of 1% lower to score 50.5% against 49.5% on a board, and this converts in Cross-IMPs to +0.0772 IMPs against -0.0772 IMPs. And for a Teams event where the stronger pair's team mates have an average grade 1% higher than their opponents, we can expect the same to happen at their table. So we expect the stronger team to gain 0.1543 IMPs per board while we expect the weaker team to lose 0.1543 IMPs per board.

Thus for a given match, you can take the average grade for the stronger team and subtract the average grade for the weaker team, and multiply this by 0.3086 times the number of boards to be played in the match, to get the expected difference between the plus score of the stronger team and the minus score of the weaker team, so this is the handicap in IMPs, provided that you add this handicap to the score of the weaker team

For a handicapped knockout event, you can give each team a handicap figure that is the average grade for that team multiplied by 0.3086 times the number of boards per match, and for any match, the team with the lower handicap starts with a number of IMPs that is the difference in handicap figures between the two teams. In other words, the stronger team must finish with a plus score



in excess of the difference in average grade multiplied by 0.1543 times the number of boards to be played in the match in order to “win” the match.

For Multiple Teams, you can give each team a handicap IMPs figure to add to their final IMPs result that is:

0.3086 times the number of boards to be played times (50 – the average grade value for the team)

For example, for a team with average grade value of 45% playing in a Multiple Teams event of 24 boards, you give them a handicap of

0.3086 times 24 times (50 – 45)

= 37 IMPs

Historic data from an annual series of handicapped teams event at Sheffield Bridge Club was used as a practical check and the results fell in line with expectation.

8. What types of Pairs events are we not considering for the National Grading Scheme?

The National Grading Scheme considers the seven most common type of Pairs events only, but it is appreciated that there are other types of Pairs events that are possible:

Knock-out Pairs

Play-with-the-expert Pairs (IMP scoring)

Pairs with Instant Match Point Scoring

These other types of Pairs event are either very rare in face-to-face bridge or impractical to score electronically, so will not be considered for the National Grading Scheme.

Knock-Out Pairs, while providing a superb competition in Bridge Club Live, is virtually unheard of in face to face bridge because it is very complicated to organise.

There is a problem with Instant Match Point Scoring in that often, when it is used, the original event was held in a different country where bidding methods vary from this country. This can result in situations such as where 3NT is played the other way up at most English tables compared to the original field. Whether it makes or not depends on the opening lead so it usually makes one way up but not the other. Thus, in England, you can get an average score nationwide of 20% or 80% for either direction. Anyway, these events are usually scored manually, so we won't concern ourselves with them.

Play-with-the-expert Pairs is also very rare other than on the Monday evening at the Brighton Congress.

So we consider just the seven main types of pairs events listed in the main text.

9. How did you get the formula for Aggregate Pairs?

In theory, it is possible to hold an Aggregate pairs event as a single winner movement, either with an arrow-switch or using a Howell movement, but as this is virtually unheard of and, as most or all scoring programs assume two-



Full Description of the National Grading Scheme

winner events, the National Grading Scheme confines itself to two-winner events.

A comparison between MP Pairs and Aggregate Pairs has been achieved by using data from Sheffield BC. The club holds two-winner MP Pairs events on Thursday evenings and Aggregate pairs events on Tuesday evenings.

The players on those two evenings are mostly the same. The standard is slightly higher on a Thursday but the range of standards is near enough the same. The range of standards is tighter than at most clubs as Sheffield BC has other evenings of play intended for other standards of players. The number of boards played on Tuesdays and Thursday is the same, 26 or 27 being normal but some movements being 24 or 30 boards. All of this means that we can reasonably compare sessions on a like-for-like basis.

The Standard Deviation of Thursday results (final percentage MPs for each pair) averaged 6.720% away from the mean of 50%

The Standard Deviation of Tuesday results averaged 1417 aggregate points away from whatever the mean happened to be for that particular direction and session.

(As a side note, Sheffield BC uses a “windfall tax” system as recommended in Farrington’s book of duplicate movements, and “windfall tax refunds” as recommended by the EBU in 1951, reference for which can also be found in Farrington. The effect of this is that it should be appreciated that the mean score in one direction may not necessarily be equal and opposite to the mean score in the other direction.)

The above statistics have taken 6 months-worth of data into account.

We consider that a pair that achieves an MP score of one Standard Deviation above average in a Thursday evening session at Sheffield BC needs to average 56.72% on each board. To achieve the same in an aggregate event of the duration of a Tuesday evening at Sheffield BC would need a pair to average 1417 points above the mean for the session, and thus about 55 aggregate points per board.

For a single board, a 56.72% score is equivalent to scoring 55 aggregate points above the mean. Thus a 100% score in MPs is equivalent to 409 aggregate points above the mean on a single board.

We need to replace the “ $p - 50$ ” bit in the MP Pairs formula.

We start by inserting $p - m$ where p = the aggregate score the pair achieve and m = the mean score for that direction. We then need to divide that by something and that something will be the conversion of 55 divided by 6.72 (which comes to 8.2), and also the number of boards played in the session, which we will call “ x ”

So the replacement is:

$$((p - m) / 8.2x)$$

where p , m and x are all described above.



10. What happens if I have been an EBU member for the last three years but played fewer than 300 boards in that time?

If you have not yet played as many as 300 boards in any three year period, your initial artificial assigned grade will remain in place for the number of boards fewer than 300 that you have played in the last three years.

If you have previously played more than 300 boards in any three year period, the effect of your initial artificial assigned grade will have disappeared, and only those boards that you have played in the last three years will contribute towards your grade. If this is the situation, you will be ineligible to appear in any lists of top graded players that may be published.

11. What happens if I stop playing bridge for three years and then restart?

If the unthinkable happens and you don't play for three years, the effect of all your previous results will have disappeared. When you resume, you will be considered as a new player except that you will be assigned an artificial grade of 50% irrespective of the prevailing artificial assigned grade for newcomers.

12. How does NGS cope with unknown players in a session?

When some players in a session are players not known to EBU, or in a few other cases with data entry errors, we cannot assign grades to the associated pairs in the process of determining the Strength of Field. For short we will call all pairs where one or both the players is unknown to EBU as an "unknown pair". For NGS purposes an EBU member partnering an unknown player is treated for that session alone as an unknown.

Originally, unknowns were simply assigned a default grade of 50, but this was found to result in over-grading, as the majority of these unknowns are infrequent club players with a relatively weak playing strength. An alternative of assigning a grade based on their actual score in the session is proposed here. In tests this has been shown to avoid over-grading, on average, the known EBU population.

In what follows define

g_m as the current known grade of each EBU member player ($m = 1, 2, \dots, M$)
 h_u as the current unknown grade of each unknown player ($u = 1, 2, \dots, U$), and
 s_u as the standardised score for each unknown.

Remember that for MP-Pairs standardised score is just actual percentage score - 50.

Denote by M and U respectively the number of member and unknowns. We wish to estimate the Strength of Field value SoF.

We first note that the SoF value is estimated from the current grade values of all $M+U$ players as

$$\text{SoF} = (\sum_M g_m + \sum_U h_u) / (M+U) \quad (1)$$

where the sums are over the guest and member players in the field respectively.



That is

$$(M+U) * \text{SoF} = \sum_M g_m + \sum_U h_u \quad (2)$$

For unknowns we are going to say that their grade is what their (standardised) score would predict. So:

$$h_u = s_u + \text{SoF for each unknown.} \quad (3)$$

Substituting (3) into (2) gives

$$(M+U) * \text{SoF} = \sum_M g_m + \sum_U s_u + U * \text{SoF} \quad (4)$$

Or

$$\text{SoF} = (\sum_M g_m + \sum_U s_u) / M \quad (5)$$

In the NGS algorithm the average over all the players of the difference between a player's current grade and their derived session grade for the session is zero.

In NGS the average of the difference between session grade and previous current grade is zero. The consequence of using formula (5) for SoF is that the derived session grade for unknown pairs is equal to their (assumed) current grade. Therefore the average of the difference between the session grades and current grades of the known players is also zero. This is what assures almost zero drift in NGS grades overall.

13. The "Aces and Eights" Issue

Further to the reply to FAQ 28 on this issue, some case histories have been examined

The overall conclusions:

1. The differences are what the NGS should give. There are no bugs or data errors involved.
2. NGS is quick to use results to estimate a difference in the grades of the players in a regular partnership.
3. Differences of up to 4 points between the grades of players in a regular partnership are likely to be due to chance, unless they have both played with several other partners who have a mature grade and who have also played with several partners.
4. Original FAQ 9 is right.
5. To misquote Harold Macmillan, "It's events results dear boy, events results."

Case 1: Andrew is an Ace Of Spades and Beth is an Eight, but they feel the difference should not be anywhere near this

However, Andrew doesn't lose ground and Beth doesn't gain ground when not partnering each other.

Our analysis proceeded as follows:

Andrew has played 3358 boards with Beth. They have a partnership grade of 59. We looked at their games with other players.

Andrew first (current grade 69):

312 boards with Simon, but Simon has played with no-one else, so his grade is underestimated by exactly the same as Andrew is over-graded whatever that is. Their partnership grade is 61, so no surprise that Simon is 52.5

288 boards with Carol who has a mature grade of 61.5, which is probably accurate (to within 3) since she has lots of partners that she's played over 300 boards with. This would suggest a



Full Description of the National Grading Scheme

grade for Andrew of 60.5+/-5 say.

144 boards with 4 other people, which by themselves would imply a grade of 65+/-5 for Andrew if the others' grades are not under-graded. We didn't check these other players. So let's guess that Andrew's grade is really 63.

Beth (current grade: 49) in addition to 3358 boards with Andrew:

408 boards with Paul (grade 59) with partnership grade of 56. Paul plays with even more partners than Carol so probably is not over-graded. In particular 436 boards with Vera - see below, with a partnership grade of 63.

384 boards with Vera (grade 63) with partnership grade of 54. Again Vera has lots of partners so probably an accurate grade.

Now we have a triangle of players Beth – Paul – Vera. This is good news for getting individual grades.

Beth + Paul = average 56,
Beth + Vera = average 54 but
Paul + Vera = average 63!!

We can solve these three equations for the three unknowns to get Beth = 47, Paul = 65, Vera = 61, all +/-6 to allow for the small number of boards.

240 boards with 7 other players, which by themselves would imply a grade of 46 for Beth if these players are not over-graded. They all have mature grades, and I see no evidence of over-grading, but it can be hard to spot.

But her partnership grade with Andrew would suggest a grade of 55 if he is really 63. Perhaps this is an example of gelling (a partnership gels when they play better together than with other partners). Because Andrew hardly plays with other players who have many other partners, whereas Beth does, if there is gelling, what NGS seems to say to itself is "well Beth's grade must be right, so it must be that Andrew is stronger than previously thought, that is to say he gets all the benefits of gelling and Beth gets none. If so, the cure is for Andrew to play with other players with a mature accurate grade, preferably in green pointed events which are weighted triple by NGS, and so bring his grade down rapidly to the low sixties (unless of course he wins these events, in which case all bets are off!)

Case 2

From Steven: *"I have a query on the grades for myself and for my wife. We play nearly all our bridge together, so you would think we'd have almost the same rankings. Currently I am ranked King on 60%, and Mary is ranked 7 on 47%. It's very hard to understand how that can happen (flattering though it is for me!)."*

Initially we thought that finally here we have a couple who have been hard done by, as Steven scored an amazing 75.25% with David back in April 2010, when NGS didn't know that David strength was around 57 but assumed he was around 50. This gave Steven a session grade of 99.85 which clearly got him off to a flying start, while Mary was getting average to poor results with players she has since stopped playing with. All this sound like the ingredients for giving Steven a falsely high grade and saddling his partner Mary with a corresponding low grade to match their partnership grade of 54.

But the truth is that NGS is just reflecting the results. Mary has played 14 sessions with other players in the last 2 years, and if we just used these to determine her grade would be around 46. Steven has played only 4 sessions with other players, and even if we recognise David's true grade for that great result of 75.25% the tentative grade for Steven for the four sessions would be 66. So inevitably NGS is going to say that Steven is a lot stronger than Mary. The gap of 13 points between their grades isn't so surprising.



Of course they haven't played enough games with other people to claim that the gap of 13 in their grades is very accurate, I'd guess that it would be fair to estimate it as 13+/- 5, say between 8 and 18.

To close the gap, either Mary has to score above par when playing with other partners or Steven has to play with other partners and score below par.

Further comments

We also know that the same kind of phenomenon occurs with exponential decay as in Bridge Club Live. However, similar situations there are far fewer, partly due to the smaller population, but also due to a smaller proportion of almost exclusive partnerships. This "Aces and Eights issue" would occur with any method of grading.

14. Volatility and decay

We were told: *"The system as implemented would appear to me to be calculating a grade that is too current; for some highly active players it will simply represent their performance over the last 2-3 weeks. This does not appear reasonable to me."*

This was asked at a time when we considered the last 1000 boards played. Our reply was: "Few players manage 40 sessions of bridge in such a short time. Some regard the rate of decay over 1000 boards as too fast and some regard it as too slow. However, we will keep this under review."

Since April 2013, we have taken into consideration the last 2000 boards played by a player.

15. Linearity of results

The following was posted on an Internet forum.

"Presumption of linearity in expected result. It's assumed that if you play in a field that is different from the norm, your expected percentage will change by the same amount regardless of what your starting percentage was. I don't think this is true. For example, suppose Frances has a rating of 68% in a national "average" field and shows up to a weak club game with a partner of comparable skill. The club game might be 10% below national average, and Frances and partner are very likely to win... but will they score 78%? Even in a weak field, it is very hard to consistently score at that level. Similarly, if you take two very weak pairs and put them in a national championship event, their final scores will be quite bad. But there will be much more luck as to who has the better score, than there would be for them in an event where the standard is weaker. My point is that when a pair is much better (or much worse) than the field their expected MP score should tail off. I don't think expected scores in excess of 75% are really reasonable or accurate regardless of the caliber of field (same could be said of scores below 25%)."

Our presumption of linearity is the foundation of the scheme, and it is a very reasonable presumption for the vast majority of events. The argument given is also quite reasonable as there becomes a limit beyond which it is difficult to imagine that linearity holds. It is our belief, though, that it would be very rare that a significantly strong player would wish to play in a significantly weak field and vice-versa, so in practice, this question of linearity will rarely be tested.

The writer then considers *"presumption of linearity in partnership caliber. Carrying a very weak partner to a good result in a mediocre event is really a very different skill from getting a good result with a comparable partner in a top event. I know a lot of people who are much better at one of these skills than the other. It doesn't seem reasonable to presume that two strong players who obtain comparable results when playing with their regular partners will necessarily do comparably well when partnering a beginner in the pro-am, yet the rating system seems to presume precisely this."*



We cannot disagree that some players perform better with one strength of partner rather than another, but the degree or extent of this is unlikely to be significant and there is no way to account for this anyway. Individuals can begin to assess which of their partnerships have greatest gelling by comparing their partnership grades with the averages of their own and their partner's grades, but we would recommend some caution as these comparisons can be affected by other unrelated factors.

16. Other interesting comments from the Internet forums

1.

"I can think of one disadvantage of this scheme: some people may not like being told that they're not very good. However, I doubt if that would cause many people to stop playing: either they'll see it as a reason to try to improve, or they'll dismiss the grading system as being inaccurate, irrelevant, and not nearly as meaningful as masterpoints."

2.

"I also feel that the effect of all recent results should be the same - ie the result from last weeks pairs should have the same effect as yesterday's pairs ... "

One of our group replied:

If you look at the degradation diagram on the Full Guide, the degrading appears as a triangle. We did discuss some other shapes. At an extreme is a rectangle. If we were to take that, then the result that drops off the left hand end will have as great an effect as the most recent event that you have played. As a halfway, we could have gone with your suggestion, which is essentially a trapezium. From the right hand side, the line goes leftwards horizontally for a distance before turning diagonally downwards. For this to have any significant difference to the triangle (straight line decay) we would have to keep the horizontal line going for some way, and then the line would turn quite steeply downwards. We were worried that this would give too much of an effect like the rectangle, and we plumped for the straight line, which is not just simpler but simpler for everyone to visualise. The exponential (or "radioactive") decay in the BCL grading scheme actually goes in the other direction in that the line goes steeply downwards to begin and gradually almost levels out.

3

"In any other sports or games, one easily understands a system on how to go about enhancing one's grading/ranking but in the case the NGS system it is not possible to do that due to its complexities and it seems to be a mathematical nightmare. Can you please advise if you have come across any other systems in any other sports/games that is as complicated as the NGS under Bridge and if so please advise which one? A vast majority of our members are over 65 and it is unlikely that they will spend time in understanding such complicated system."

One of our group replied:

I am not a skilled mathematician at all, nor a grading expert, but the simple answer here is, if you beat your expected or "PAR" score, shown on the session file, your grade will go up. If you don't, it will go down. If you play with a stronger partner, your "PAR" required will be higher than if you play with a lower partner. What else does one need to know? Rather like Golf!

As a parallel I have been driving for over 40 years without knowing how the car works, yet I can still drive reasonably well.

4

"Could you explain please how Swiss Pairs VPs are converted to MPs. In a recent County Congress Swiss Pairs we came 19th out of 114 pairs scoring 120 VPs out of 200 which I make



60%, yet our MP score for NGS somehow became only 52.84%. This percentage seems very low for a pair finishing in the top sixth of the field."

One of our group replied:

It uses your match point percentage, before it's converted into VPs. I don't think 53% is that low. If this was a 70-board pairs event (which is essentially how the NGS treats it), you might expect a winning score around 60% and 53% might well be a top-sixth score. Look at the results from the London Easter Championship Pairs for an example.

5

"Just a comment.

Every time someone takes out a lease/hire purchase/repayment mortgage/annuity an extremely complicated formula is used to determine the payments.

Now I say "extremely" because it looks like a load of gobbledegook - however it is mathematically justifiable, it works, and it is accepted without much debate by many people.

The formulae used for these rating calculations are a little more arbitrary, but they achieve the aim of giving a plausible numeric value for performance. I expect that a very large number of people will just accept it. There will always be a few who want to twiddle and/or object, but to what purpose, even after their twiddle there will be someone else who wants to twiddle some more."

17. Concerns over grading Teams events

We were asked:

My bridge club is considering whether to allow the results of multiple Teams of Four events to be reflected in the NGS rankings of participants. We have some concerns about the method used for rescoring teams events to allow them to be included in the NGS rankings.

Teams events are heavily influenced by the scoring method, the current state of the match and the perceived foibles of the other 7 players involved in each match. Surely these latter 2 factors mean that it is only valid to create rankings based on the results within each match, and that comparisons with other matches playing the same board(s) are not legitimate?

Reading the guide tells me that for NGS purposes Teams events are rescored on a cross IMP basis, meaning that scores are compared with other tables with whom there is no perceived or actual competition at the table. But in certain circumstances players would behave differently depending on whether cross IMP scoring is used (as per NGS) or not (as per the actual event at the table).

I have read the guide 'The EBU grading system and Teams events' and I presume that the statement 'and yes there are rare exceptions to this rule' reflects the considerations I have outlined above. But are these exceptions really that rare?

An alternative would be to produce rankings based only on the results within each head to head match. I recognise that these would then be based on limited competition, across just 2 teams for each match; however, where the event involves multiple matches, such as multiple Teams of Four events, this would tend to balance out across the whole event.

In summary, our concern is that a Teams event is scored for NGS purposes on a significantly different basis than that at the event itself. Isn't this a little like taking a head to head match at the Ryder cup and rescoring it as stroke play?

I am sure much thought has been given to the scheme, and that other people have raised



similar questions to these along the way. Hopefully therefore you will have an answer to our concerns

We replied:

You are right that this has been discussed at length; the approach we use is the best compromise and we believe is fair.

You are right that the comparison to make is between what the two pairs do on at one table on one board in a Team of Four match (in a multiple teams event) versus what the same pairs would do if the board was in a cross-IMP scored pairs event. Well, generally they will do the same. Occasionally, they will think "oh I'll do X because obviously our teammates would have opened a strong no trump, whereas in a cross-IMP pairs event they'll think "I'll do Y because most of the room is playing a weak NT." Even more occasionally this will result in a significantly different score on the board. Doubtless there are some other examples where having an idea about what is likely to have happened to your teammates affects your bids or play and results in a different score (compared to a cross-IMP Pairs strategy). How rare? That is a matter of opinion, watching boards played by strong players on vu-graph in BBO, I'd say it was a very minor factor in players' strategy, and as a less strong player myself I know it is a very small factor in my strategy!

There is a second factor, which affects Pairs matches as much as Teams matches, namely the state of the match. If you are oodles ahead, play safe. A real factor, I'm sure and less rare I guess, but still a comparatively small factor in a match since it can only affect the last board or two of a match.

There is data however. A team's IMP score over all the boards of a session is approximately the sum of the two pairs' cross-IMP scores, so trying to maximise one will pretty much increase the other by the same amount.

Let me give a real example, the winners of a club multiple teams event (28 boards, 8 teams, 4 board matches) scored a total of +60 IMPs, the NS pair scored +43 Cross-IMPs and the EW pair 15.3 Cross-IMPs. You might say that the discrepancy between the sum of Cross-IMPs (58.3) and the total IMPs (60) was due to one or the other pair taking account of the factors above. Maybe or maybe it is just random fluctuation. But the big factor affecting NGS is the difference between 43 Cross-IMPs for NS and the 30 IMPs we'd credit them with if we split the 60 IMPs between them. It was easy to see that this was because NS bid that thin game and that EW made that phantom sacrifice, because we have other boards to compare with, not simply a subjective post-mortem.

These are much larger factors than the rare bias due to strategy based on guessing how our teammates have performed, as so I am convinced that including these games in NGS in this way is correct.

18. Research arising from the NGS

The following paper was written by NGS Project Leader, Mike Christie, in October 2013.

Variability in Duplicate Pairs Bridge events

Summary

This paper presents the results of an analysis of variability using the data from half a million boards played by EBU affiliated clubs and at EBU affiliated events. Some boards played are flatter than others; many pairs scoring the same number of points. This paper defines a quantitative measure of flatness, and reports on the distribution of flatness.



It also looks at the variability of a pair's overall match point percentage score for a session and the factors that influence it. The analysis shows that the EBU's National Grading Scheme (NGS) grades are an accurate predictor of the non-random component of variability.

Finally, the variability of scores when using IMPs scoring for pairs is examined and the appropriate conversion factors for comparing IMP and match point scores are discussed.

Statistical definitions

This paper analyses various aspects of variability when playing in duplicate pairs bridge events. Variability will usually be measured using the standard deviation (sd) of some quantity about its average (mean) value. Often we will use the variance (var) which is the square of the sd, as the components of the variability we are analysing can be split into different parts, whose variances are additive when looking at the variability of the whole.

Certain data which have been collected or calculated can be regarded as samples from a large population of possible values; such data are called random variables, and the mean, sd and var of the population from which they are samples are "population" statistics.

Detailed statistical or mathematical justifications of the formulae used are not given here. The reader will either have to take them on trust or derive them for himself/herself.

In games scored according to match points, pair get a score for each board depending on the number of other pairs they have beaten or drawn with. We are mostly concerned with this score expressed as a percentage of the maximum possible score, the Match Point Percentage (MPP).

The EBU database of sessions.

Since April 2010 the results all sessions played by EBU affiliated organisations have been uploaded to an EBU database. The data are used by the EBU's National Grading System (NGS) to estimate the playing strength of all 50,000 plus members. Most, but not all, of these session results include the sort of information that would be on a traveller for each board, including the match points score by each pair on each board. This has enabled us to compute the MPP variance on each board and the sample value of the session MPP variance for each session. The scoring systems do not record what adjustment method was used when different boards have been played different numbers of times, and even if they did one might not trust this to be reported accurately! Therefore for our statistical analysis we have used only those sessions which met the following criteria:

- a) The sessions were match point scored pairs sessions. (See later for other types of sessions.)
- b) Over two-thirds of the pairs were EBU members and had a NGS grade (most sessions met this criterion).
- c) There were at least 7 tables in the session.
- d) Board by board data were available for the session.



- e) Each board in the session was played the same number of times, so that no corrections to the MPP scores will have been made (about one-third of the sessions did not meet this criterion). This restriction is needed since the reporting does not make clear exactly what method of correction was used.

Over the 12 month period August 2013 to July 2014, there were 17,759 such sessions in the database, containing a total of 461,894 boards, nearly half a million. We could have extended our analysis back in time but this seemed to be enough for a reliable statistical analysis!

Match point scoring in a large field of pairs.

First of all imagine a very large number of tables playing the same board, and consider the match point percentage score for the NS pairs. It will range between 0 and 100 with a mean of 50. First pretend that every pair gets a different match point score; of course that's not possible with a large field so suppose pairs getting the same result are ranked at random. Then the Match Point Percentages (MPP) will be uniformly distributed between 0 and 100. This distribution has a variance of $2500/3 \approx 833.3$. Note that this will be the case irrespective of whether all the pairs are equally strong or whether they have a range of abilities. It can be shown that ties (pairs getting the same match point score) always reduce the variation of the match points. So 2500/3 is the maximum variance of MPP for a single board. Of course if a board is very "flat", for example where everyone will bid and make 3NT exactly, the variance will 0. If the outcome depends on a two-way finesse for which there are no clues, half the field will make and half will go one down. They will score 75% and 25% respectively and for every pair the deviation from the mean of 50 will be ± 25 , so the sd will be 25 and the variance will be 625. If the contract is hard to make or depends on a lucky guess, and just 10% make the contract (the good pairs and/or the lucky pairs), they'll score 95% and the rest 45%. The variance in this case is 225.

Essentially, the amount by which the variance of the match point scores is less than the maximum measures how flat the board is.

In practice a single board is played by a limited number of pairs. In a typical club session it is often played as many times as there are rounds (most commonly 8 or 12). If a board is played n times, it can be shown that the maximum variance of MPP is:

$$\text{maxvar}(n) = (2500/3) * (n+1)/(n-1)$$

which happens if there are no ties for match points. As before, any ties reduce the variance.

Over the course of a session, some boards will be flatter than others, and if we take the variance of MPP on each board, divide this by the theoretical maximum variance for that board, and finally take the average of this over all the boards of the session, we will have the average flatness of boards in the session from a match points perspective. This will be denoted by F .

For a single board the MPP variance, which we will denote by MPPvar is given by the formula:

$$\text{MPPvar} = F * \text{maxvar}(n)$$



Analysis of flatness

The average value of F over all these boards was 0.898 (near enough to be called 0.9). This means that the overall variance of MPP on a single board for a large field would be $0.9 \times (2500/3) = 750$, and so the sd would be $\sqrt{750} =$ approx 27.4%.

Perhaps of more interest when playing a typical session is to work out the sd of a single pair's MPP score in a session, using the formula:

$$sd = \sqrt{F \times (2500/3) \times (n+1)/(n-1) \times (1/b)}$$

For a typical session of 24 boards with 3 board rounds, and each board played 8 times, this formula gives an sd of 6.34%. Roughly speaking this means that for about 2/3 of the match-pointed pairs sessions you play you will score within 6% of your par score and the remaining third will be split between over 6% above and more than 6% below par.

Only a small percentage of the half million boards had a flatness factor of less than 0.8 as the following table shows. The 0.3% in the right hand column represents the 3 boards per 1000 where all pairs scored the same.

Flatness	0.9-1.0	0.8-0.9	0.7-0.8	0.6-0.7	0.5-0.6	0.4-0.5	0.3-0.4	0.2-0.3	0.1-0.2	0-0.1
% of boards in this range	68.7%	19.5%	6.0%	2.4%	1.4%	0.7%	0.9%	0.1%	0.0%	0.3%

All of the sessions we have analysed were graded by NGS, and so we know the average grade of the players in each session. As you might expect the stronger the sessions the fewer wild results are scored and so the "flatter" the boards on average, where flatness is as defined by the factor F. This is not a big effect though as the following table shows. Even the sessions with strong players seem to have a flatness just 4% lower than the norm.

Average grade of players in session	% of EBU members in this grade	Total number of boards	F value
less than 47	40.7%	48,658	0.912 = 0.9 * 1.013
47-53	36.3%	340,482	0.898 = 0.9 * 0.998
53-57	14.6%	67,517	0.885 = 0.9 * 0.983
over 57	8.5%	5,237	0.864 = 0.9 * 0.960

Session match points percentage

Let us ignore all the complications of different movements, and possible arrow switches. So just think of the all NS pairs playing in a full Mitchell movement with



no arrow switches, where all the pairs play all the boards; we are going to compare the total NS match points (as a %). If all the pairs are of equal strength each pair would expect to score 50% with a variance of $F \cdot \text{maxvar}(n)/b$, where n is the number of boards each pair plays in the session.

If the pairs have differing strengths, there is an additional component to the session MPP variability, due to the variability of the NS pairs' actual strength. Here we measure actual strength as the expected mean MPP the pair would expect to get against average opposition. The variability of NS pairs' strength, for that session, is the variance of the strength of all the NS pairs playing in that session (PSvar). Of course players' actual strengths cannot be known precisely, only estimated, so PSvar cannot be known exactly, and we will come back to this later. The mean of the session MPP scores for all the NS pairs is still 50% and it can be shown that the variance of these session MPP scores is the board variance plus the pair strength variance, that is:

$$\text{sessionMPPvar} = F \cdot \text{maxvar}(n)/b + \text{PSvar} \quad (1)$$

It is important to note that this is what statisticians call a "population" variance. Here this means that if we imagine lots of similar sessions with the same values for b , n , F , PSvar then the variance of all the session MPP scores will be as given by the population variance. For a single session the "sample" variance of the set of actual scores will be a random variable. Statistical theory says this will have a chi-square distribution, with $p-1$ degrees of freedom where p is the number of pairs' scores being compared. This prevents us from taking the observed sample value of sessionMPPvar from a single session and using equation (1) to determine PSvar.

Allowing for the actual movements of real sessions.

Most real club sessions do not have a full Mitchell movement and often involve arrow switches; nevertheless equation (1) is a good approximation. There are likely to be bigger fluctuations when there are few tables in the session. More significantly sometimes some boards are played more often than other boards, and some adjustments to the MPP scores have to be made to compensate for this. The best method for doing this is to use something called the Neuberg formula, but this is not always used by the scoring system and even then there are variations in how it may be applied. It is for this reason that such sessions were excluded from the analysis.

Session MPP variance and the accuracy of NGS grades

As mentioned above, the true playing strength of a pair is not known, but we can use the average NGS grade of the two partners in the pair as an estimate.

Pair playing strength = average grade of the two partners + pair-grading-error

Actually what matter is a pairs playing strength relative to the average strength of the session, and in this case the mean of pair-grading-error over all the pairs in a session is zero.



Now we can split the session variance of pair playing strength, PSvar in equation (1), into the sum of two parts; namely the session variance of the average grade of each pair and the variance of the pair-grading error. We'll denote these by Gvar and Evar respectively. For each of the sessions we have analysed we recorded Gvar. We also computed the term $F \cdot \max \text{var}(n)/b$ in equation (1) for each session. We'll denote this by Rvar, the variability in a pair's score due to the random factors when playing a session of bridge. Then equation (1) becomes:

$$\text{SessionMPPvar} = \text{Rvar} + \text{Gvar} + \text{Evar} \quad (2)$$

As mentioned before this is a "population" variance. We also recorded the "sample" score variance of the actual session MPP for the pairs in the session, denoted by Svar.

Initially, naively, the author had hoped to get an estimate of the accuracy of grades in a single session from equation (2) from the equation:

$$\text{Estimate of Evar} = \text{Svar} - \text{Rvar} - \text{Gvar} \quad (3)$$

since each of the terms on the right can be computed for each of the analysed sessions. This is naïve since the right hand side is a random variable and it can be shown that the standard deviation of the right hand side is typically around 20%MP, whereas we hope that the grading error is much smaller than this, of the order of 1 or 2%. However, we can use equation (3) for all 17,759 sessions and get an estimate of the average value of Evar rather than that for a single session.

The average values over all these sessions are:

Svar	57.83
Rvar	39.43
Gvar	17.68

Hence the average value of Evar is 0.72 with a confidence interval +/- 0.3. Taking square roots, the standard deviation of "pair-grading error" is somewhat less than 1% (between 0.65% and 1.0 %). This means that the average grade of the two partners in a pair will measure their true playing strength relative to the other players in the session with an error of less than 1% on average.

This is not the same as saying an individual's grade is on average accurate to within 1%, as an error in one player's grade may be compensated by an opposite error in their partner's grade, and furthermore the players in a session may be over or under graded compared to the average EBU membership. These factors contribute noticeably for players with few partners if those partners also have few partners, and also affect clubs which have sessions with very few players who play at other clubs or in sessions of mixed strength. These errors are much harder to quantify, but analysis carried out in the early days of NGS suggest these errors increase the average grade error to between 2 and 3%.

In passing, we also computed for each pair in a session how much they have scored above or below their "par" score based on their grades; indeed this is at the core of NGS grade calculations. Since grades should predict a pair's session score the variance of all pairs score above/below par (Avar) should be lower than Svar. Again these are random variables, so this will not always be



so for a single session, but the mean value of Avar over many sessions should be $Svar - Gvar$. It is reassuring to note that for our analysed sessions the average value of Avar was 40.24 and the average of $Svar - Gvar$ was 40.15; the difference is well within the sample error margin.

Sessions scored by IMPs

A small percentage of the pairs sessions submitted to the EBU are scored by IMPs, usually cross-IMPs but some using Butler-IMPs. These have been analysed to measure the variability in cross-IMPs on a single board. Again the variance is used, this time of the cross-IMPs for all the pairs who played a board (Xvar).

This is of more general interest as it can be shown that Xvar is exactly half the variance of the IMP score on a board played between two teams of four. This variance (or XRvar, defined below) and its corresponding standard deviation have been the subject of a number of earlier studies and values of the IMP sd of between 6 and 7 have been reported, with the consensus nearer to 6 than 7.

Using similar criteria as before, but taking a three year period from August 2011 to July 2014, a total of 502 sessions comprising a total of 13585 boards were analysed.

In an analogous fashion to the analysis of MPP, the population variance of the cross-IMPs for a single board, played at n tables is:

$$Xvar = XRvar * n / (n-1) + XPSvar \quad (4)$$

Here XRvar is the cross-IMP variance for a large field of equally strong pairs and XPSvar is the variance in playing strength of the pairs who played the board, measured in cross-IMPs per board. From the data, we got a value for Xvar for each board and for XPSvar we used a value derived from the NGS grades of the players. This enabled us to compute a value of XRvar for each board.

The overall average for XRvar was 20.61 (+/- 0.4). This corresponds to an IMP-sd of 6.42 (+/-0.06) which is certainly consistent with previously reported results.

These sessions are not representative of the full EBU membership. Many members do not play any IMP scored pairs events, and the average grade of those playing in these sessions was noticeably above average; the average grade of players in these sessions was nearly 55. The results have been analysed by the average grade of the players playing.

Average grade of players in session	Total number of boards	XRvar value
less than 49	299	20.5 (+/-3)
49-51	1986	19.7 (+/-1)
51-53	1703	19.8 (+/-1)
53-55	2784	20.7 (+/-1)
55-57	3283	21.5 (+/-1)
57-59	2557	21.3 (+/-1)



over 59	973	18.9 (+/-1.5)
---------	-----	---------------

Because of the comparatively small sample sizes, the XRvar value has quite broad confidence limits, shown in the table as +/- values. Nevertheless, the tendency for very strong players to have a slightly lower than average variability can be seen from those sessions where the average grade was over 59, as has been reported in other case studies. The increased variability for sessions with an average grade between 55 and 59 is a mystery. However, it happens that most of the results in this grade range come from Pairs sessions using Cross-IMP scoring from a single club, the Young Chelsea Bridge Club in London. It seems these sessions are somewhat more variable than the norm! If this club's results are excluded, the overall average of XRvar reduces to 20.0 (which corresponds to an IMP-sd of 6.32 IMPs).

Recently, the NGS has been extended to those Teams events where the cross-IMP score of each pair can be provided. In future this will give us a larger sample of cross-IMP sessions, with perhaps a wider range of players.

As an aside, we also analysed those sessions played with Butler IMP scoring. Given the relatively small sample size, this adds little to our analysis, but did support the empirical conversion rate that says 1 cross-IMP is worth 1.2 Butler-IMPs.

Comparing match point results with (cross-) IMP results

One of the reasons for this analysis was to provide a detailed justification for the way NGS compares MPP values with cross-IMP values. We say that a pair with a grade of 50+p will expect to score (50+p)% in a match-pointed session against average opponents and want to be able to say they will expect to score +q cross-IMPs per board in an IMP scored session against average opponents using a proportionality factor between p and q. From inception, NGS has used a factor $X = 6.48\%$ as equivalent to 1 cross-IMP per board; so $q = p/X$.

In this analysis we shall choose a value for X, by comparing the sd of the random factors in the two sorts of scoring, in each case assuming there was a large field of players. Explicitly, we suppose that:

$$XRvar = MPPvar/X^2 = F*(2500/3)/X^2.$$

Using the values of $F = 0.9$ and $XRvar = 20.0$ gives a value of $X = 6.12$ rather than the value of 6.48 we have been assuming.

However, most EBU graded sessions aren't played in sessions with "a large field of players". On average the number of times a board is played is around 9 times. Given that the number of times a board is played affects match point scoring and cross-IMP scoring in different ways this gives a different formula for X, namely:

$$X^2 = [F*(2500/3)*(n+1)/(n-1)] / [XRvar*n/(n-1)] = [F*(2500/3)/XRvar]*(n+1)/n$$

Using the EBU average value of $n=9$ gives a value of $X = 6.45$. It could be argued that NGS should take account of the number of times a board is



Full Description of the National Grading Scheme

played to give more weight to sessions where the boards are played more often, rather than adjusting the IMP to MP conversion factor X in this way. (That is impractical for NGS, as often results are reported in a way that does not identify how often boards have been played.)

Perhaps one should focus on the way the strongest players perform and use the values $F = 0.864$ and $XRvar = 19.0$; this gives a value for $X = 6.16$.

It is clear that one should not be too dogmatic about this conversion factor. Using round figures always seems good, so the following values are suggested:

flatness factor, $F = 0.9$

cross IMP variance (against equal opponents), $XRvar = 20.0$

hence IMP variance = 40.0

MPP to cross-IMP ratio = 6.12%, equivalent to 1 cross-IMP per board,

MPP to Butler-IMP ratio = 5.10%, equivalent to 1 Butler-IMP per board.