Human Errors in Qualitative Respiratory Protective Equipment Fit Testing: A Study of Real-World Fit Testers

Karen M. Long[1,*], Nathalie Mai[1], Michael Williams[1]

1. Cranfield University, Cranfield Defence and Security, Bedford UK.

* Corresponding author email: karen.long@face-fit.co.uk

ABSTRACT

Background. Fit testing is an essential part of any respiratory protective device program and provides a method for assessing the effectiveness of the face-to-facepiece seal of a tight-fitting facepiece which helps to assure wearer protection. Although qualitative fit test methods are assumed to be simple and easy to use, it remains critical that the methodology is applied in accordance with the protocols set out in guidance to ensure that a correct ‘pass’ or ‘fail’ result is assigned. An incorrectly assigned pass result increases the risk to the wearer of exposure to hazardous respiratory substances in the workplace, putting their health or life in danger. Fit testing stakeholders, manufacturers and regulatory bodies have raised concerns about the quality of fit testing conducted in the UK workplace, but there have been no published research studies in this area to date.

Methods. This article presents results from a study into the errors made by real-world qualitative fit testers, using the data gathered during Fit2Fit practical assessments conducted in the UK over a ten-year period from 2009 to 2019.

Results. When application of the qualitative fit test method was measured against the HSE INDG479 protocol, assessment records indicate that fit testers made a median of 4 errors, and a high of 29 errors. Fit2Fit accredited fit testers appear to make half as many errors as non-accredited fit testers, yet the scheme remains voluntary in the UK. The sources of errors were analysed in detail and revealed that the most common errors were made in instructing the wearer, donning and fit checking RPE, verifying the nebuliser function, and assuring the wearer’s taste detection.

Conclusion. Results from this study suggest that errors made by real-world fit testers are common, and steps to mitigate the assignment of false pass fit test results arising from these errors are needed if wearer health is to be better protected.

Keywords: Qualitative fit test, QLFT, FFP3 respirator, half mask respirator, Fit2Fit, competency.

INTRODUCTION

Fit testing is an essential part of any effective respiratory protective device (RPD) program and is a crucial link in the chain of assuring worker protection. Tight-fitting facepieces, which include filtering (disposable) half masks, reusable half masks and full-face masks, must be suitably matched to a wearer’s individual facial features to minimise leakage via the face seal. The UK Approved Codes of Practice supporting the Control of Substances Hazardous to Health (COSHH) Regulations 2002 require that all reasonable steps be taken to prevent or reduce exposure to substances hazardous to health (Health and Safety Executive, 2005). Where RPD is selected on the basis of a risk assessment, it must provide adequate protection against the hazard and, in order to be suitable for the wearer, the initial selection of tight-fitting facepieces should include a fit test (Health and Safety Executive, 2013). Whilst fit testing remains a key element of UK RPD guidance, other national regulatory bodies have been slower to act, with only a handful of countries formally implementing the requirements of the International Standard ISO 16975-3:2017 or setting any local recommendations for fit testing of RPD (British Standards Institution, 2017).

Guidance on fit testing of RPD facepieces for the UK workplace is set out in the Health and Safety Executive (HSE) industry guidance document HSE INDG479 (Health and Safety Executive, 2019). This guidance is supported through the British Safety Industry Federation (BSIF) Fit2Fit Companions which provide complementary 'best practice' guidance in a series of documents which accompany the main three fit test methods: qualitative, quantitative (ambient particle counting) and quantitative (controlled negative pressure) (British Safety Industry Federation, 2023b, 2024a, 2024b). Quantitative fit test (QNFT) methods employ the use of an instrument to assess the amount of face seal leakage, and produce a numerical value known as a fit factor (FF). Qualitative fit test (QLFT) methods provide a simple 'pass' or 'fail' result and are restricted in their use to filtering (disposable) and reusable half masks, however these methods are in widespread use in the UK workplace due to their relatively low cost and perceived ease of use.

QLFT methods use the wearer's sense of taste (or less commonly, smell) to identify face seal leakage of a bitter- or sweet-tasting challenge agent and thus to assign a fit test outcome. The original QLFT method was developed almost 40 years ago, and little has changed in the design of the rudimentary equipment or methodology in the intervening period (Marsh, 1984; Mullins, Danisch and Johnston, 1995). QLFT equipment requires manual operation and management, and thus the method relies heavily upon the skills of the human fit test operator. During both the sensitivity and fit test procedures, aerosolisation of the challenge agent is achieved through actuation of a hand-operated jet nebuliser, and containment / management of the aerosol is achieved through use of a test hood. Proper use of this equipment and application of the test protocol are fundamental to the principles of the QLFT method, and errors made by human operators introduce the potential for those principles to be undermined. Any process that involves a human operator has the potential to be affected by human errors, and errors which result in an unreliable fit testing outcome are of serious concern for the millions of workers that rely on the protection afforded by tight-fitting RPD. The importance of properly assigning a 'pass' or 'fail' fit test outcome cannot be underestimated; RPD facepieces may be the last (and sometimes only) line of defence against a respiratory hazard, and a wearer that is assigned a pass result in error (a false pass result) potentially risks increased exposure in the workplace. False pass results can occur in all fit testing methods and arise when a wearer with an unsuitable (leaking) facepiece erroneously passes a fit test. A false pass fit test result offers the wearer false confidence that a facepiece is safe to use in the workplace, when in truth it is not, potentially putting their health or life at risk.

Sources of errors in the QLFT method are complex, and review of the current literature in this area identified three key sources of error in this method: 1) statistical errors inherent in the QLFT method, 2) errors arising from the subjective nature of the method, and 3) human errors made in real-world application of the method. Whilst the first two of these areas are largely well understood, further research is needed, especially in understanding the wide range of statistical errors reported and the problems of subjectivity associated with this method (Coffey, Zhuang and Campbell, 1998; Coffey *et al.*, 2002; Frost, Mogridge and Roff, 2015). The biggest gap identified in the literature was the paucity of research into human errors in QLFT. Almost all research into the QLFT methodology has thus far been conducted under laboratory conditions, with those studies which have been conducted in the workplace focusing on the outcomes of workplace protection factors studies or respirator type. Numerous commentators have expressed concerns and cited anecdotal evidence about human errors in real-world application of the QLFT method, yet to date there have been no published studies aiming to enhance the understanding of human errors in QLFT (Campbell *et al.*, 2005; Clayton and Vaughan, 2005; McKay, 2016). Development of the BSIF RPE Fit Test Providers 'Fit2Fit' Accreditation Scheme (Fit2Fit scheme) acknowledged that a lack of competency exists in the UK workplace, with the potential to cause serious harm to the health of the wearer (British Safety Industry Federation, 2023a). Indeed, the BS ISO Standard 16975-3:2017 (British Standards Institution, 2017) highlights that the reliability of the validated QLFT method depends upon the proper application of the accepted test protocol by the fit tester:

> "Any variation from the procedure specified below can invalidate the results, including changes in solution concentration, how the bulb is squeezed, the number of squeezes, and the size of the fit-test hood."

Under UK guidance, fit testing should only be conducted by a competent person (Health and Safety Executive, 2019). A competent person as defined by the British Standard BS ISO 16975-3:2017 is, "person with suitable and sufficient experience and with practical and theoretical knowledge of fit-test methods".

Although not mandatory, HSE INDG479 guidance cites the Fit2Fit scheme as one way in which fit testers can demonstrate their competence in fit testing. The Fit2Fit scheme has a set of well-defined assessment criteria which closely reference HSE INDG479, providing a consistent and controlled evaluation mechanism. The primary purpose of the records (assessment papers) gathered during the Fit2Fit assessment process is to enable a judgement to be made of a fit tester's competence. Since recording is made meticulously against traceable criteria, these assessment papers also provide a detailed insight into those areas of the fit testing process which the fit tester conducts correctly, and those in which errors are made. This thereby offers a unique opportunity for secondary research into the human errors displayed by real-world qualitative fit testers in the UK.

The process of the Fit2Fit assessment which results in Fit2Fit accreditation is shown in Figure 1. Fit testers may apply for assessment in any (or all) of the three of the fit testing methods which are described in HSE INDG479.

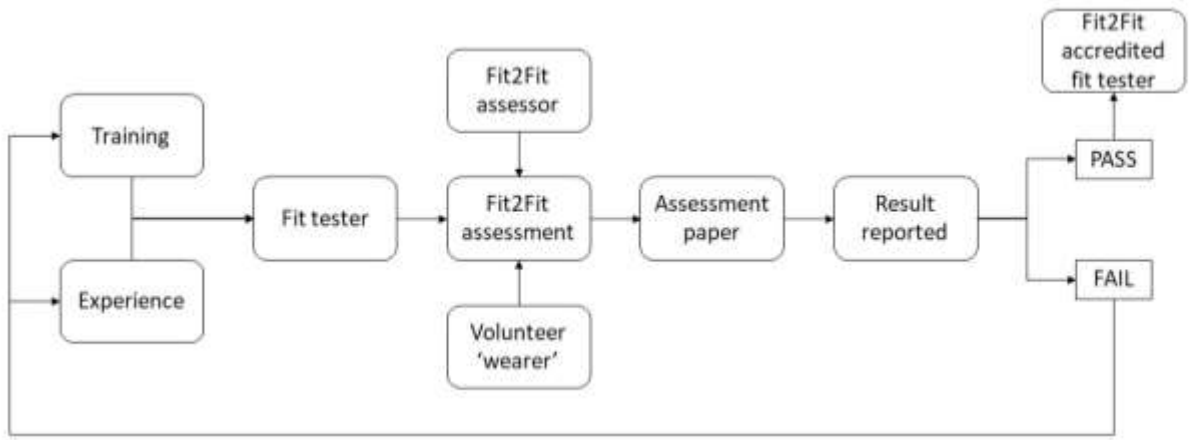

Figure 1. Schematic of the process leading to Fit2Fit accreditation. Fit2Fit accreditation is awarded by the British Safety Industry Federation (BSIF).

The Fit2Fit assessment process is designed to examine the knowledge, understanding and practical skills of RPD fit testers through a theory examination and practical assessment. Successful candidates are awarded the 'Fit2Fit accreditation' and are required to re-sit the practical assessment every 3-years, and the theory examination every 6 years to maintain their accreditation status. The theory examination is designed to assess a fit tester's knowledge of relevant guidance, and the practical assessment is designed to assess a fit tester's practical skill and ability to apply the method. Since this research study is concerned with investigating human errors in the application of the method, analysis was undertaken for the practical assessment only. Further studies to assess the fit tester's knowledge using the theory examination may be useful in providing information to authors of guidance and standards, or fit test training courses in the QLFT method.

The importance of research in understanding human errors in QLFT cannot be underestimated; the protection afforded by tight-fitting RPD facepieces is dependent upon an effective facial seal, and therefore human errors which lead to a poor-fitting facepiece being used in the workplace may have very grave consequences indeed. This study is the first to use the primary data gathered about the performance of UK fit testers during Fit2Fit assessments as a research tool for addressing three important research questions: 1) how frequently do real-world fit testers make errors in application of the QLFT method; 2) what are the main sources of human errors in QLFT; and 3) what are the implications and recommendations, if any, for use of the QLFT method in the UK workplace? Research into human errors in QLFT, along with an understanding of where and how errors are made is crucial if false pass results are to be mitigated and the workplace safety of this fit test method assured. This in turn is essential for ensuring that wearers of tight-fitting facepieces are afforded the protection they expect in the workplace.

METHODS

Data collection

The primary data (assessment papers) used in this research were recorded by nine different Fit2Fit assessors during Fit2Fit practical assessments conducted between 2009 and 2019 at numerous different locations around the UK. A complete dataset of 656 QLFT assessment papers, for 396 different fit testers (subjects), was made available for secondary research at BSIF offices, along with supporting protocols and procedures. This dataset provided a record of the observations of over 800 hours of practical QLFT for a wide range of different filtering (disposable) and reusable half mask respirators, using a pool of volunteer wearers who were not required to go into live hazardous workplace environments. Wearers participating in these assessments had received no formal training in RPD and were asked to read the pre-assessment documentation provided by Fit2Fit to provide consent for participation. During the practical assessment, the fit tester was required to prepare and conduct one or more fit tests for a wearer using a range of facepieces applicable to the fit test method. Each practical assessment was observed and marked by a Fit2Fit assessor who used an assessment paper to record compliance or deviation by the fit tester (an error). Assessment was conducted against a set of prescribed assessment criteria, which reference the HSE INDG479 protocol and Fit2Fit accreditation was subsequently awarded on the basis of the assessor's judgement.

For the purposes of this study, each of the 656 available assessment papers was investigated using a methodical paper-by-paper secondary analysis process. Data was anonymised and transferred from each hard-copy assessment paper to an electronic spreadsheet for secondary inquiry and statistical analysis. Various versions of the assessment paper were found to have been used over the 10-year period, with additional sub-criteria and varying recording fields. Integration of all versions was necessary to allow analysis of the entire dataset, and therefore sub-criteria were merged and recording fields were dummy-coded. Of the total dataset available, 17 assessment papers were found to be either incomplete or partially illegible and were therefore removed, leaving a total of 639 assessment papers included in this study.

Scoring and assessment criteria

A Fit2Fit assessment paper equips the Fit2Fit assessor with a set of 49 assessment criteria and three recoding options against each: 'Fully covered', 'Partially covered', and 'Not covered'. Since the purpose of this research was to investigate the prevalence and sources of human errors, these options were streamlined and dummy-coded into a binary scoring system. This scoring system used '0' to denote a 'Fully covered' outcome (indicating compliance with the HSE INDG479 protocol), and '1' to denote a 'Partially covered' or 'Not covered' outcome (an 'error', indicating deviation from HSE INDG479 guidance).

To provide clear terms of reference, each of the 49 Fit2Fit assessment criteria were referenced by linking to the corresponding criteria in both the Fit2Fit syllabus and HSE INDG479 guidance document. The data were analysed both at individual-criterion level and in groups according to the Fit2Fit assessment paper section headings, which facilitated analysis of specific criteria and wider trends. Data were handled in spreadsheet software (Microsoft Excel, Microsoft Office 365) and symmetry of the data was investigated using a skewness test. Appropriate descriptive statistics were used to examine the data, and non-parametric statistical analysis was performed using IBM SPSS software, version 26 (IBM Corporation, Armonk, NY). A Mann-Whitney U test was used to compare the error scores for sub-groups according to accreditation status of non-accredited and Fit2Fit accredited subjects (significance level = 0.05).

RESULTS

The outcomes of this study reveal that of the 639 (N) Fit2Fit QLFT assessment papers investigated in this study, 30.0% (n = 192) indicated that the fit tester failed to meet the required standard to award Fit2Fit accreditation. In this analysis, N represents the total number of assessment papers studied (639), while n represents the number of assessment papers in a subgroup of the total. Table I displays the descriptive statistics for observations recorded at Fit2Fit QLFT assessments and provides additional data about accreditation outcome. The mean (M) number of errors recorded in assessments for fit testers that were judged to have failed the Fit2Fit assessment was 11.2 (n = 192), compared to a mean of 3.2 (n = 447) for those that passed the Fit2Fit assessment.

Table I. Descriptive statistics by assessment outcome of errors recorded in Fit2Fit QLFT practical assessment paper for the period 2009 – 2019.

Fit2Fit outcome	Recorded error scores		
	M	***Mdn***	**Range**
All	5.6	4	0 – 29
Pass[a]	3.2	2	0 – 22
Fail[b]	11.2	11	1 – 29

Notes: N = 639. [a]n = 447. [b]n = 192.

The frequency of errors recorded for fit testers attending Fit2Fit QLFT practical assessments for the period 2009 – 2019 is illustrated by the histogram in Figure 2. Possible error scores ranged from zero to 49, with a higher score indicating a greater number of errors and a higher deviation from the HSE INDG479 protocol. Error scores in the data set showed a minimum score of zero errors and a maximum score of 29 errors. The mean score (M = 5.6) was greater than the median (Mdn = 4), indicating that the distribution had a longer tail towards higher scores. This positive skew was illustrated by the shape of the frequency histogram in Figure 2 and confirmed by a skewness statistic that was positive (*skew* = 1.34), indicating that the data were not normally distributed.

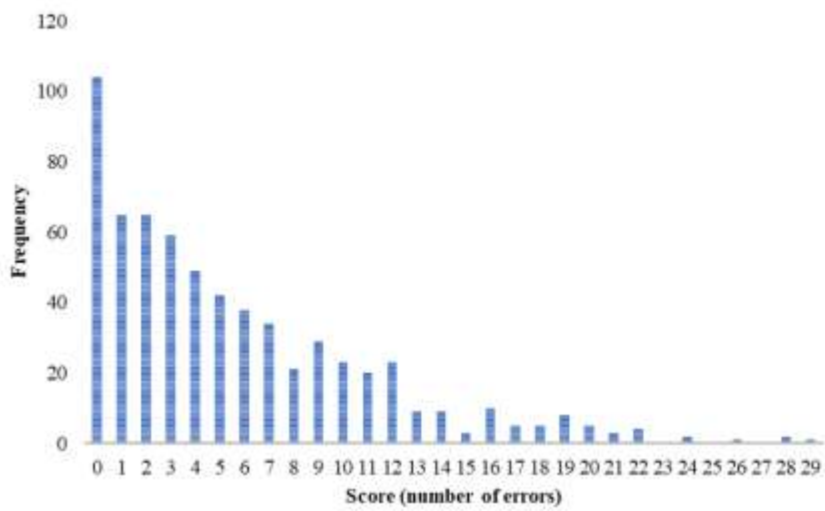

Figure 2. Frequency of errors recorded in Fit2Fit QLFT practical assessment papers for the period 2009 – 2019 (*N* = 639; maximum possible number of errors = 49).

These data show that one or more errors were recorded for fit testers in 83.7% (n = 535) of assessments, and an interquartile range of 1 to 9, shows that between 1 and 9 errors were recorded for fit testers in half of all assessments. The data presented in Table II further indicate that five or more errors were recorded in almost half of all assessments (46.3%, n = 296), ten or more errors in one fifth of assessments (20.7%, n = 132), and twenty or more errors in 2.7% (n = 17) of assessments.

Table II. Grouped frequency of errors recorded in Fit2Fit QLFT practical assessment papers for the period 2009 – 2019 (*N* = 639; maximum possible number of errors = 49).

Number of errors recorded	Frequency	% (*n*)
0	104	16.3
>0	535	83.7
>5	296	46.3
>10	132	20.7
>20	17	2.7

Table III provides a summary of the descriptive statistics for each year of the study period (2009 – 2019). These results indicate a trend towards a broader range of error scores over this period of study, with the five highest mean and median scores in the latest five years.

Table III. Descriptive statistics by year of errors recorded in Fit2Fit QLFT practical assessments papers for the period 2009 – 2019 (N = 639).

Year	Recorded error scores			
	n	***M***	***Mdn***	**Range**
2009	20	2.8	3	0 - 10
2010	10	1.6	1	0 - 5
2011	32	5.1	3	0 - 29
2012	42	4.5	3	0 - 18
2013	54	5.0	4	0 - 19
2014	63	4.4	3	0 - 15
2015	100	6.4	6	0 - 19
2016	98	5.7	4	0 - 28
2017	105	6.5	5	0 - 21
2018	104	6.4	4	0 - 26
2019	11	7.0	5	0 - 20

When the year of assessment (time variable) was regressed on the number of errors observed (Figure 3), the resulting unstandardised regression coefficient was $b = 0.391$. The value of this coefficient showed that, for every one-year increase since the start of the scheme, there was an increase of 0.391 observed errors as illustrated in Figure 3. The R^2 value for the regression model was 0.03, indicating that year of assessment accounted for only about 3% of the variance in error score.

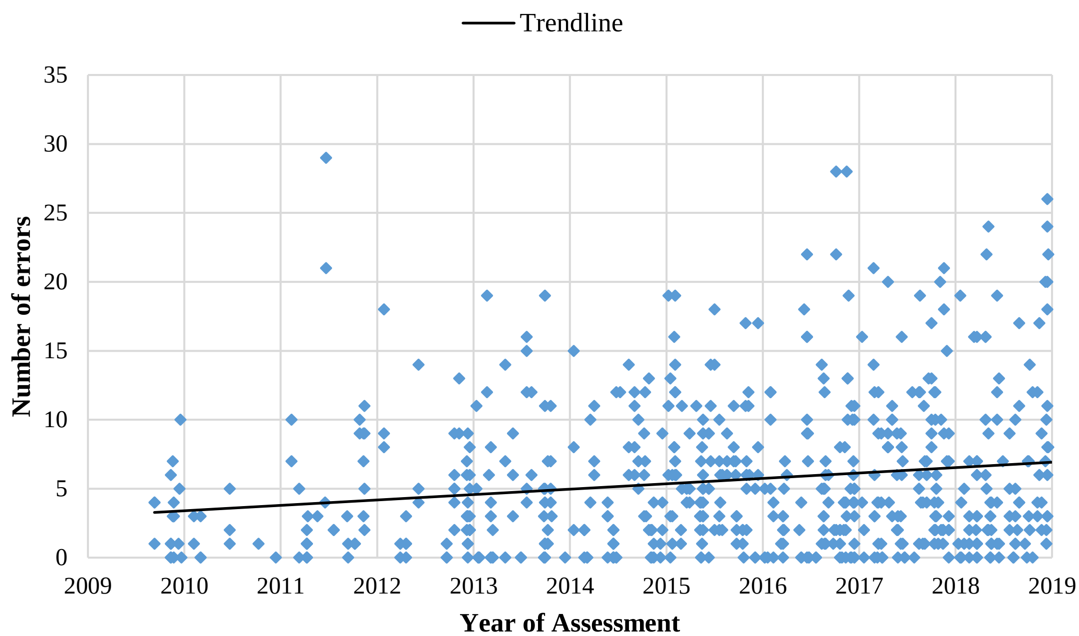

Figure 3. Regression analysis for year of assessment against number of errors recorded in Fit2Fit QLFT practical assessments papers for the period 2009 – 2019.

Data gathered in this study also provided the opportunity to examine the descriptive statistics according to 'accreditation status', these results are summarised in Table IV. Of the 639 assessment papers investigated in this study, approximately three quarters (77.3%, *n* = 494) were conducted for non-accredited fit testers, and a quarter (22.7%, *n* = 145) were for those already Fit2Fit accredited and returning for a three-yearly re-assessment.

Table IV. Descriptive statistics by accreditation status of errors recorded in Fit2Fit QLFT practical assessments papers for the period 2009 – 2019 (N = 639).

Accreditation status	Recorded error scores		
	M	*Mdn*	Range
Non-accredited[a]	6.4	5	0 - 29
Fit2Fit accredited[b]	3.0	2	0 - 22
All	5.6	4	0 - 29

[a]*n* = 494, [b]*n* = 145.

The range of error scores recorded in assessments for the group of Fit2Fit accredited fit testers (0 - 22) was found to be similar to that for the group of non-accredited fit testers (0 - 29). However, the mean and median error scores for the group of Fit2Fit accredited fit testers (*M* = 3.0, *Mdn* = 2) was lower than that for the non-accredited group (*M* = 6.4, *Mdn* = 5) suggesting that Fit2Fit accredited fit testers achieve higher compliance with the HSE INDG479 protocol. A Mann-Whitney test indicated a significant difference in the distributions between the two groups (Mann–Whitney U = 22 311.5, $n1 + n2 = 639$, $p < 0.05$ two-tailed). The result is significant ($p < 0.05$), suggesting that the group of Fit2Fit accredited fit testers make fewer errors compared to non-accredited fit testers.

This study was also designed to examine the sources of errors made by fit testers in Fit2Fit assessments. Table V shows the percentage of total errors for each of the Fit2Fit sections. The highest number of errors were observed during the sections covering: *Conduct fit test* (24.7%), *Recognise wearer related factors prior to carrying out a fit test* (15.5%), and *Conduct sensitivity test* (13.4%).

Table V. Frequency of errors by section recorded in Fit2Fit QLFT practical assessment papers for the period 2009 – 2019 (N = 3584; % (n) = percentage (number) of all errors).

Section	Fit2Fit section heading	% (*n*)
1	Ensure suitability of facilities	1.5 (55)
2	Preliminary actions prior to performing a qualitative fit test	3.3 (119)
3	Brief wearer on purpose of fit testing and test exercises	7.0 (252)
4	Examine the face piece for suitability prior to fit testing	11.2 (402)
5	Confirms the wearer has presented themselves as required	8.7 (311)
6	Conduct sensitivity test	13.4 (481)
7	Recognises wearer related factors prior to a fit test	15.5 (555)
8	Conduct fit test	24.7 (887)
9.1	Response if tasted (i.e., a fail)	11.2 (403)
9.2	Response if nothing tasted (i.e., a Pass)	3.2 (113)
10	Post Fit Test Procedures	0.2 (6)

Detailed analysis of data for each of the 49 Fit2Fit criteria reveals additional information about the sources of errors made by fit testers. Table VI shows the 'top ten' errors recorded in assessment papers investigated in this study. These criteria are ranked in order from highest to lowest prevalence with higher scores indicating a higher deviation from the HSE INDG79 protocol. The most common errors made by fit testers, each recorded in almost one third of assessment papers were: *Ensures fit check is performed* (32.2%, *n* = 206), and *Demonstrates instruction on correct method of donning an RPD including a fit check* (31.3%, *n* = 200). Approximately one quarter of all assessment papers indicated that the fit tester made errors in each of the criteria: *Verify correct function of nebuliser during the test* (25.2%, n = 161) and *Questions the wearer about the use of other PPE* (24.4%, *n* = 156). The least common errors recorded by assessment papers investigated in this study were: *Clean Hood* (0.5%, *n* = 3), *Dispose of Solutions* (0.3%, *n* = 2), and *Rinse & clean nebulisers* (0.2%, *n* = 1).

Table VI. Top ten sources of error recorded in Fit2Fit QLFT practical assessment papers for the period 2009 – 2019 (N = 639; % (n) = percentage (number) of assessment papers that recorded an error).

Fit2Fit Assessment Criteria	% (*n*)
Ensures fit check is performed	32.2 (206)
Demonstrates correct method of donning an RPD including a fit check	31.3 (200)
Verify correct function of nebuliser during the test	25.2 (161)
Questions the wearer about the use of other PPE.	24.4 (156)
Advises wearer to rinse mouth and clean face after sensitivity test	24.3 (155)
Repeat test sensitivity Test	23.2 (148)
Verify palette is clear of taste	21.0 (134)
Explains the purpose of fit testing	20.3 (130)
Explains the fit test exercises and procedure	19.1 (122)
Administer solution as per sensitivity level	18.2 (116)

When Fit2Fit outcome (binary Pass or Fail) was compared with the assessors' observations (binary compliance or deviation) against each of the 49 assessment criteria, the strongest relationship was found for Administer top ups during fit test ($\phi = 0.508$, $p < 0.001$). The correlation between Fit2Fit outcome and Administer fit test solution as per sensitivity level was also strongly statistically significant ($\phi = 0.474$, $p < 0.001$), as was the correlation between Fit2Fit outcome and Verify correct functioning of nebuliser (ϕ =

0.451, p < 0.001). Whilst this finding indicates that these areas are most strongly correlated to Fit2Fit outcome, this may also suggest that these areas are judged most critically by assessors.

The primary data used for this study were recorded by nine different Fit2Fit assessors over the ten-year period. Table VII shows the recorded error scores by assessor, where identifying data have been removed to assure anonymity.

Table VII. Descriptive statistics by Fit2Fit assessor of errors recorded in Fit2Fit QLFT practical assessment papers for the period 2009 – 2019 (N = 639; % Fail = percentage of assessment papers that recorded a fail overall result).

Fit2Fit Assessor	Recorded error scores				
	N	***M***	***Mdn***	**Range**	**% Fail**
1	36	8.1	8	0 - 19	50.0
2	236	6.2	4.5	0 - 29	26.3
3	138	8.3	7	0 - 29	44.2
4	122	3.4	2	0 - 22	23.0
5	15	5.3	5	0 - 10	26.7
6	3	7.3	7	3 - 12	66.7
7	25	1.8	1	0 - 5	4.0
8	32	2.6	2	0 - 12	21.9
9	32	2.7	1	0 - 17	25.0
All	**639**	**5.6**	**4**	**0 - 29**	**30.0**

Error scores recorded by the nine assessors show considerable variation, with mean error values ranging from 1.8 (assessor 7) to 8.3 (assessor 3). Similarly, the number of Fit2Fit failures judged by the nine assessors varied widely with a range of 4.0% (assessor 7) to 66.7% (assessor 6). The large variation for assessors 6 and 7 may be a function of a small sampling population since the number of assessments conducted by these assessors was low (n = 3, n = 25) compared to the total (N = 639). By contrast, the error scores for assessors 2, 3 and 4, who together have conducted 77.6% of all assessments, show greater consistency (M = 3.4 – 8.3, % Fail = 23.0 – 44.2).

DISCUSSION

Since the widespread requirement for fit testing under the COSHH Regulations 2002 there has been a huge uptake of fit testing in the UK workplace. Once the remit of just a handful of experts, RPD Fit Testing is now conducted by in-house fit testers as well as a multitude of external providers. False passes are by far the most significant from the perspective of wearer-health and therefore the need for information about the prevalence and sources and of human errors in QLFT is of paramount importance.

The results of the study into the frequency with which errors are made indicates that human errors are common and widespread, suggesting that the majority of fit testers do not apply the QLFT method strictly in accordance with the HSE INDG479 protocol. Less than one fifth of all assessment papers indicated that the fit tester applied the protocol wholly correctly (making zero errors), with over four-fifths recording errors against one or more of the Fit2Fit assessment criteria, and the worst-performing recording errors in 29 out of a possible 49 criteria. These results are all the more surprising given that the Fit2Fit scheme is voluntary and hence subjects in this study may have come from a population that deem themselves ‘good enough’ to apply for examination and value the ‘best practice’ approach supported by Fit2Fit. In addition, since the subjects were aware of being observed and marked during the assessment, their behavior may have altered. This may have resulted in the fit tester to being on ‘best behavior’, thus exhibiting fewer errors than they would do in their normal activity. This could suggest that human errors made in the application of QLFT in the wider UK workplace may be more prevalent than those observed in this study.

It might be expected that the number of errors made during Fit2Fit assessments would have decreased over time with greater awareness, availability of training resources and improved practical guidance from HSE and Fit2Fit. Analysis of the trend in data over the 10-year period in this study do not appear to support this expectation however, and instead show a trend towards higher deviation from the HSE protocol with increasing raw error scores and range over the period. However, due to the lower number of assessments conducted in earlier years, it is difficult to know whether this is a true effect, or a function of a lower sampling population. Another factor to consider here is that fit testers presenting for assessment in earlier years were more likely to have been key stakeholders, including regulatory bodies, RPD manufacturers and full-time fit test providers. These groups could be expected to have a better knowledge of guidance, more practical experience, and better preparation, all of which may have contributed to improved performance and lower error scores. The true picture of the performance of real-world fit testers may therefore be closer to the data found for later years, where the scheme had reached those in the wider UK workplace.

The results show that the individual ability of fit testers varies considerably and those that are Fit2Fit accredited appear to make half as many errors, and hence achieve higher compliance with the HSE INDG479 protocol than their non-accredited counterparts. These results support the Fit2Fit mission statement “to improve the respiratory health of those wearing tight fitting Respiratory Protective Equipment through being face fit tested by a Fit2Fit Accredited provider”. Although Fit2Fit accredited fit testers appear to deviate less from the HSE protocol than non-accredited fit testers, this study was not designed to examine the reasons for this apparent difference in performance. Fit2Fit accredited fit testers in this study were returning for a 3-year reaccreditation, and therefore had the benefit of attending a previous accreditation plus at least 3 years’ experience. The post-assessment feedback provided by the assessor and scheme administrator provides valuable information which the more diligent candidate may use in improving their performance in future assessments, hence demonstrating fewer errors. Of the 639 assessment papers analysed in this study, 243 papers were recorded for candidates that had returned for a first, second or third re-accreditation or re-attempt. Future research using this dataset could be used to reveal trends in error scores and types of errors made during repeat assessments. Analysis of these trends alongside information about further study could be used to indicate to what extent any improvement in competence is achieved following re-training.

The second important area investigated by this study was an examination of the sources of errors which was evaluated through analysis of Fit2Fit sections and individual assessment criteria. The overlap in data relating to Fit2Fit ‘pass’ or ‘fail’ outcomes provides evidence to suggest that the source of errors is as, or more, important than the number of errors. Analysis shows that the best-performing fit testers in the ‘fail’ group made only 2 errors, whereas the worst-performing fit testers in the ‘pass’ group made 22 errors. One note of caution here: since a pass / fail outcome is awarded on the basis of the assessor’s judgement alone, then there is the potential for assessor bias and further analysis including additional assessor data would be required to answer this question. If indeed assessors judge some errors to be more critical than others, then the scheme would be more robust and would benefit from a clearer set of guidelines for assessors, perhaps indicating which criteria are to be judged most critically.

The secondary data used in this study were recorded by nine different Fit2Fit assessors. The variation in error scores recorded by some assessors may be indicative of differences in harshness of judgement which could therefore skew the data used in this study. In addition, cognitive biases originating from their own skills, experience, perceptions, or points of view, may impact an assessor’s subjective view of performance. Such biases are difficult to overcome in practice, however a systematic approach to ensure that assessors receive regular training, monitoring and joint observation could all help in setting and maintaining a consistent standard for Fit2Fit. Since assessors typically operate in different geographical locations, this may result in the data for some assessors to be skewed by industry / sector type. Further analysis of this secondary dataset could examine and compare the prevalence and sources of errors made by fit testers from within specific sector types. This information may be useful to regulators in focussing awareness and training campaigns.

It is perhaps not surprising that most errors were found for the section *Conduct fit test*, since this section is the crux of the assessment process, and accounts for 10 of the total 49 individual assessment criteria. Indeed, of the ten Fit2Fit sections, those which involve the set-up, use and application of fit test equipment accounted for more than half of all errors recorded during Fit2Fit assessments. This is further emphasised by the data which shows that almost a third of assessment papers recorded that the fit tester failed to verify the function of the nebuliser at key points during the procedure. Analysis of the individual assessment criteria reveal that errors are made across a wide variety of different sections. Surprisingly, four of the top ten most common errors relate to the instruction given by the fit tester to the wearer in donning / fit checking the mask or explaining the purpose of the fit test and exercises. At first sight, these areas may not appear to directly influence the ability of the method to determine the fit of a respirator, however the wearer's understanding of the principles of this method is crucial to ensuring an effective fit test. Since this method is subjective and relies wholly upon the wearer's cooperation and honesty in taste detection, their comprehension of the test protocol, their correct reporting of detection of taste, and the function of the test exercises cannot be underestimated, and errors in these areas may have critical consequences. This is further highlighted by assessor comments against some of these criteria that noted "a complete lack of explanation of the basic tenet of taste-test QLFT", and even "told wearer that the aim is NOT to taste the solution!".

The ability of the fit tester to correctly administer the taste solution, and the wearer to correctly report taste are key to the reliability of the QLFT method. The fact that four of the most common errors relate to delivery of the taste aerosol and management of the wearer's palate is therefore of concern. The QLFT method relies solely on the wearer's taste response in determining face seal leakage, and errors which affect the wearer's ability to detect the challenge agent have the potential to lead to false pass results and may have serious consequences for the wearer's health. The frequency of errors identified in this study related to taste detection corroborates findings from other researchers, who recommend incorporating a placebo control during sensitivity testing (McKay and Davies, 2000). This modification may improve the reliability of the test by eliminating unreliable responses that may lead to false pass results. In addition, failure to administer the taste aerosol correctly has the potential to limit the ability of the fit test to identify face seal leakage, allowing a poor fitting respirator to pass the fit test. Assuring proper generation of the taste aerosol by verification of nebuliser function is fundamental to ensuring that the respirator seal is properly challenged, and this area must be emphasised during fit tester training programs.

This study contributes to the empirical understanding of human errors in qualitative fit testing and has broad implications for workplace fit testing programs. Human errors were found to be widespread amongst the UK fit testers presenting for Fit2Fit accreditation and this highlights the very real potential that many of the qualitative fit tests conducted in the wider UK workplace are error prone. This casts doubt on the reliability of results obtained from workplace QLFT, which could have significant consequences for the selection of respiratory protective devices based on these outcomes.

The prevalence of human errors found in this study highlight the importance of fit test operator competence in this fit testing method and hence the need to ensure that fit testing programs are robustly supervised. The lower error rate observed in the study amongst the group of Fit2Fit accredited individuals suggests that the UK Fit2Fit accreditation scheme provides a good indication of fit tester competence amongst this group. Given the risks associated with fit testing errors of falsely assigning a poor fitting RPD to a wearer, this study would support mandatory regular competence assessment for fit testers.

By extending the understanding of QLFT to include the impacts of human errors, this study opens some new avenues for research in this field. The first direction of research would be to complement the outcomes of this study through additional real-world observational studies to further examine the extent of human errors amongst the wider population of fit testers. Secondly, a deeper understanding of the drivers for human errors could be gained through studies which examine the behavior of fit testers, including their learning and motivational patterns. Research into awareness and understanding of human errors amongst fit testers would provide additional information about the risks in workplace fit testing and support the call for further guidance in this area. Such studies may also seek to better understand the effects of increased time and work pressure upon a fit tester's role, and how these additional demands impact the fit tester's behavior.

CONCLUSIONS

These results indicate that a significant proportion of Fit2Fit assessment papers recorded that errors were made by the fit tester in application of the QLFT method. Fit2Fit accredited fit testers appear to make fewer errors than non-accredited fit testers, but the results indicate no improvement in performance over the ten-year period of this study. The most common individual errors made during assessments investigated in this study were in ensuring proper donning and fit checking RPD, and improvements in the fit tester's instruction to the wearer are needed in this area. The indication that areas involving the set-up, use and application of fit test equipment accounted for more than half of all errors recorded during Fit2Fit assessments highlights a potentially serious problem and these areas require further investigation to ascertain the impacts of such errors. The purpose of this study was to investigate the prevalence and sources of human errors to gain a complete understanding of errors in QLFT and ultimately to improve the workplace safety of this method. Whilst many of the errors made by fit testers have the potential to undermine the test protocol, further research is needed to evaluate the impact on QLFT outcome and hence the likelihood of these errors leading to false passes in workplace fit testing programs.

REFERENCES

British Safety Industry Federation (2023a). *Fit2Fit candidate handbook*. Available at: https://www.fit2fit.org/resources (Accessed: 1 October 2020).

British Safety Industry Federation (2023b). *Fit2Fit Companion to HSE INDG479 guidance on respiratory protective equipment (RPE) fit testing, qualitative method*. 2021 v2. Available at: https://www.fit2fit.org/resources.

British Safety Industry Federation (2024a). *Fit2Fit Companion to HSE INDG479 guidance on respiratory protective equipment (RPE) fit testing, quantitative ambient particle counting method*. March 2024. Available at: https://www.fit2fit.org/resources.

British Safety Industry Federation (2024b). *Fit2Fit Companion to HSE INDG479 guidance on respiratory protective equipment (RPE) fit testing, quantitative controlled negative pressure method*. March 2024. Available at: https://www.fit2fit.org/resources.

British Standards Institution (2017). Respiratory protective devices. Selection, use and maintenance. Part 3: Fit-testing procedures. BS ISO 16975-3:2017. London: British Standards Institution.

Campbell D.; Doney B.; Groce D.; Greskevitch M.; Syamlal G.; Coffey C. C.; Bang K. M. (2005). Respirator fit testing practices in the U.S. J. Int. Soc. Respir. Prot. 22(1) 11–16.

Clayton M., Vaughan N. (2005). Fit for purpose? The role of fit testing in respiratory protection. Ann. Occ. Hyg. 49(7) 545–548. doi: 10.1093/annhyg/mei046.

Coffey C.C., Lawrence R.B., Zhuang Z., Campbell D.L., Jensen P.A., Myers W.R. (2002). Comparison of five methods for fit-testing N95 filtering-facepiece respirators. Appl. Occ. Env. Hyg. 17(10p 723–730. doi: 10.1080/10473220290107002.

Coffey, C. C., Zhuang, Z., Campbell, D. L. (1998). Evaluation of the Bitrex™ Qualitative Fit Test Method Using N95 Filtering-Facepiece Respirators. J. Int. Soc. Respir. Prot. 16(1) 48–55.

Frost, S., Mogridge, R., Roff, M. (2015). *RR1029 Review of fit test pass criteria for filtering facepieces Class 3 (FFP3) respirators.* Health and Safety Executive. Available at: https://www.hse.gov.uk/research/rrpdf/rr1029.pdf.

Health and Safety Executive (2005). *L5 Control of substances hazardous to health. The Control of Substances Hazardous to Health Regulations 2002 (as amended).* Approved Code of Practice and guidance L5 (Fifth edition). HSE Books. Available at: http://www.hse.gov.uk/pubns/priced/l5.pdf.

Health and Safety Executive (2013). *HSG53 Respiratory protective equipment at work. A practical guide.* (Fourth edition). HSE Books. Available at: http://www.hse.gov.uk/pubns/priced/hsg53.pdf.

Health and Safety Executive (2019). *INDG479 Guidance on respiratory protective equipment (RPE) fit testing.* HSE Books. Available at: https://www.hse.gov.uk/pubns/indg479.pdf.

Marsh, J. L. (1984). Evaluation of saccharin qualitative fitting test for respirators. Am. Ind. Hyg. Assoc. J. 45(6) 371–376. doi: 10.1080/15298668491399965.

McKay, R. T. (2016). Respiratory Protection Newsletter. Available at: www.fit2fit.org/resources (Accessed: 16 February 2021).

McKay, R. T. and Davies, E. (2000). Capability of respirator wearers to detect aerosolized qualitative fit test agents (sweetener and Bitrex) with known fixed leaks. Appl. Occ. Env. Hyg. 15(6) 479–484. doi: 10.1080/104732200301269.

Mullins, H., Danisch, S., Johnston, A. (1995). Development of a new qualitative test for fit testing respirators. Am. Ind. Hyg. Assoc. J. 56(11) 1068–1073. doi: 10.1080/15428119591016278.