International Council for Building Research Studies and Documentation

CIB

Working Commission W60
The Performance Concept in Building

Performance test methods and the interpretation of results

Co-ordinator
Dr E J Gibson
BRE Princes Risborough Laboratory
Princes Risborough
Bucks HP17 9PX
England

December 1979

PERFORMANCE TEST METHODS AND THE INTERPRETATION OF RESULTS

CONTENTS

will be suitable for the purpose yet economical, and the requirement that manufacturers should not be called upon to produce a very wide range of products.

Paper 6 has been discussed in depth within the Commission and as a result the Commission recommends as follows:

> 'Performance requirements and the results of tests of performance in relation to those requirements are best expressed in relation to a number of banded levels. Of the various methods available for designating bands, the method which is open to least misinterpretation and provides maximum flexibility is based upon the use of letters. The normal range of performance requirements for average buildings in temperate zones should be expressed in three bands, the scale reading (K)LMN(O) or, if five bands are chosen, (J)KLMNO(P.) The extreme bands (K and O, or J and P) are intended for situations in which more or less severe requirements are appropriate – eg for arctic or tropical conditions'.

It should be noted that the bands have not been labelled A B C or 1 2 3 but L M N. The reasons for this are two-fold; one is to make it easier to superimpose further bands for situations in which requirements are more extreme, and the other is to reduce the impression that one grade is better than another rather than more appropriate, for this could lead to specifying qualities higher than are necessary for a particular purpose.

Most of the attention devoted to performance testing has been applied to building products and components, but the concept is equally applicable to the spaces of the complete building, and to elements containing a number of different components within the fabric. (See The performance concept and its terminology*, sub-section on The nature of performance.)

Generally, performance tests on buildings in situ are very expensive to carry out, and such tests are of most value when (a) providing feedback to designers, or (b) used as a means of checking performance when defects or contractual problems arise. Because they are only conducted after components have been installed in the building, they are probably not of great help in component development.

Although there have been discussions in W60 about applying the concept at the level of complete buildings, the Commission as a whole does not consider that developments have reached the stage when it is possible to publish any conclusions or advice. Indeed, we have had to advise others that the state of knowledge is not yet adequate to form the basis for the production of Standards. However, various methods have been used from time to time for verifying the performance of buildings by in situ tests and this may be a topic for further study by the Commission.

The Commission hopes that these papers will provide useful guidance to those working to implement the performance concept.

E J GIBSON
Co-ordinator, CIB Commission W60

---

*CIB W60 Publication, December 1975

DEVELOPMENT OF PERFOMANCE TESTING METHODS – SOME CONSIDERATIONS
by Georg Christensen
Danish Building Research Institute

CONSIDERATIONS AT THE PLANNING STAGE

When planning the development of a Performance Test Method (PTM) it is important to ensure the purpose of the testing is understood by all those engaged in the development work. It should be made clear that the purpose of the work is not to ascertain a perhaps rather arbitrarily decided property, but to estimate the behaviour of a specific part of the building (eg a component) when it is exposed to conditions simulating the in-use situation. Estimation of the behaviour in use can be made by means of testing in the laboratory, or by field testing in the partly or completely finished building.

It is important to realise that test development work should only be undertaken if there is an obvious need for a test method in order to permit correct evaluation. In this connection it should be borne in mind that quite often fairly simple calculations can be used instead of a test method for assessment purposes. This is especially the case if a good correlation between theoretical calculations and behaviour in use is established.

For example this is the situation in the evaluation of the structural performance of traditional reinforced concrete structures.

In the early stages of the development work it must also be very carefully examined – and re-examined – whether the results that will be obtained will be suitable for an evaluation of the behaviour of the product under in-use conditions – in other words whether the test method has a good validity.

ALTERNATIVE TESTING METHODS FOR QUALITY CONTROL

A PTM may be so complicated and costly that it is unrealistic to expect it to be used for quality control purposes, eg in a factory for building components. If in an early stage of the the development work it is evident that a PTM will be very complicated and expensive to use it may be worthwhile spending some time looking for an alternative test method which specifically for quality control purposes can simulate in-use conditions in simple tests (not necessarily performance tests and preferably non-destructive) which give results which can be shown to be critical indicators of performance.

In order to make it feasible to use a simpler, alternative, method, it is, however, very important to establish a well-documented relationship between the results from the original PTM and the alternative method. In this connection it should be mentioned that in the case where a PTM is replaced by a more traditional testing method (non-PTM) this latter should never be referred to as a PTM.

The decision to replace a complex PTM with a more simple method will often depend on the purpose of the testing. The order of decreasing complexity will generally be:

a PTM for research purposes
b PTM for general development purposes
c Quality control test method for production control.

It should be underlined that a simple quality control test method can only be used for a particular product and only in cases where the relationship between this method and a

PTM has been clearly demonstrated. Any major modification in the product will mean that a new relationship between the two methods has to be established.

Example:

When developing a testing method simulating the onslaught of rain on an external wall component or element, the method should ideally contain not only a water spray arrangement but also provision for creating static and pulsating air pressure as well as air movement perpendicular and parallel to the surface. The control of air movement in different directions during a test with driving rain is possible in a few laboratories working especially in the research field. For practical development work, however, most driving rain apparatus is, for economic reasons, only equipped for creating static and pulsating pressure and wind velocities perpendicular to the vertical surface. Experience over a number of years has shown that this simplification of the PTM is permissible when testing ordinary window and wall components in the laboratory. For quality control purposes, however, this testing procedure can often be further simplified, since the pulsating pressure can be replaced by a constant pressure. This simplification is quite permissible where it has been demonstrated that a relationship exists between the results utilising a pulsating and a static pressure for **the particular component which is going to be tested in a quality control scheme**. Any major technical modification in the window component requires that a new relationship must be demonstrated between the results from the PTM and the quality control testing method.

Example:

The durability of plywood should preferably be checked by means of a PTM simulating ambient temperature and humidity conditions. However, such conditions vary considerably and a testing programme simulating a wide variety of exposures would be very expensive to carry out. Instead a boil test can be considered as suitable for a wide range of applications within building since experience has shown a reasonable relationship between test results and actual durability for the adhesive types in current use.

WILL THE TESTING METHOD BE INDEPENDENT OF THE MATERIAL OR CONSTRUCTION METHOD?

It is desirable that a PTM in principle is independent of the material or construction tested under in use conditions. However, it is difficult to maintain this principle in all cases. This is mainly due to lack of knowledge in the testing field and theoretical work should consequently be undertaken in order to improve this knowledge. However, until more knowledge is developed it is often reasonable to carry out 'performance testing' for different types of materials.

Example

For the evaluation of the weather resistance of polymeric materials a number of accelerated testing methods are available. The choice of method(s) must, however, depend on the type of polymeric material involved since different mechanisms are responsible for the degradation process in the different types of material.

HOW ACCURATE NEED A PTM BE?

A testing method should of course not require more technical equipment and qualified staff than necessary. It will very often be a choice of accuracy against cost. If in doubt the rule should be to choose the simple test rather than the more complicated and accurate test since it is very seldom that high accuracy is required. However, simplification should not go so far that the method does not give a reasonable simulation of the in-use conditions.

Example:

The impact strength of a partition will depend on the elastic properties of the partition (total as well as partial) and the elastic properties of the impact body. Nevertheless a sandbag testing method where the elastic properties of the impact body are neglected is used in most countries because an 'elastic approach' is quite complicated and hardly gives significantly different results.

PTM AND GRADING OF PRODUCT QUALITY

Whenever possible an attempt should be made to develop methods which make it possible to use the test result for grading of product quality. This means that in principle the testing methods which only give information such as pass/fail should be avoided. Where the test method makes grading of test results possible, such information will make the use of banded quality levels feasible eg as used in general lists of functional requirements and product information files. When test results are expressed in quantitative terms it will also be possible to make adjustments of the banded levels in the light of experience and new information without loss of data based on past tests.

Example:

Testing of the impact strength of building components is often made with a sand-bag falling from a defined height, and the test result will give information of the passed/failed type. It would be fairly easy to refine such a method by using various fall heights with predetermined intervals. The results from the testing could then be used to differentiate between different qualities in relation to impact strength.

DOES THE PTM HAVE A HIGH DEGREE OF VALIDITY?

In the first place a testing method should only be named a 'Performance Testing Method' if it has a clear relationship to in-use conditions. If it gives results which are in accordance with experience from behaviour in practice, the PTM can be regarded as valid. It follows that the test methods with the highest validity will usually simulate accurately in-use conditions.

Example:

The composition of the materials in a PVC floor-covering has an important effect on its abrasion properties. It is rather tempting to make a chemical analysis and then make statements about the quality. In some cases this will give reasonably correct results but it should be realised that such a method is not a 'Performance Testing Method'. Here an abrasion test must be considered as being more valid since it is to a great extent representing the in-use conditions. A chemical analysis could be made for quality control purposes – but such testing is not considered as a PTM.

ACCURACY, REPEATABILITY AND REPRODUCIBILITY

A PTM must be of such a nature that the testing equipment and procedure used is capable of giving results which have sufficient accuracy for the purpose. It must also be possible to repeat the test on the same sample or a similar test specimen and obtain the same results within the required limits of accuracy (repeatability). In the same way it should also be possible to obtain basically the same results from tests performed in different laboratories (reproducibility). These statements apply of course for all testing methods and not only for performance testing methods, but often the simulation of the in-use conditions makes it more difficult to reach a high degree of repeatability and reproducibility, together with a high degree of accuracy.

Example:

It should be determined whether the operation of the testing equipment by the laboratory technicians has an influence on the results. If this is the case, more accurate operational instructions must be worked out in order to secure a better repeatability as well as a better reproducibility.

WILL THE RESULT FROM THE PTM BE UNDERSTOOD CORRECTLY?

A PTM must preferably be of a nature which makes it immediately evident that there is a reasonable link to the in-use conditions. Further it is important that the person evaluating the result of the test method is not misled by the name of a test method.

Example:

It has been stated that polystyrene is shown by a certain test method to be 'self-extinguishing'. This has caused much confusion and also surprise when fires did occur involving such materials. Another example is the use of the term 'water resistance' testing, where the consumer is misled into believing that the meaning of the term is absolute – which it certainly is not in the majority of cases.

RELATIONSHIP INFORMATION/COST

A PTM must be looked upon from a testing point of view in exactly the same way as any other testing method. This means that when it comes to problems in connection with sampling, repeatability and reproducibility, the general rules of statistics should be used. Such considerations often lead to expensive testing programmes when more simple methods would be more realistic. In such cases, simple cost/benefit considerations may show the most realistic testing procedure.

Basically PTM s ought to be rather simple. However, a performance testing method should be used even if it is expensive, if it gives sufficient important information. It will often be the case that a sophisticated and very realistic PTM is necessary at an initial stage of a research study, but later it can be replaced by simpler methods which have been shown to give results with sufficient validity.

Example:

Rather simple spread-of-flame testing methods have been developed on the basis of full-scale fire tests. Instead of performing very realistic but also very expensive tests on whole buildings (or parts of buildings), it is sufficient to test in a small apparatus. (It can be argued that a true relationship remains to be proved for this particular example).

FINAL REMARKS

Before spending much time is developing a performance testing method considerable effort should be devoted to analysing the in-use conditions so as to make sure that the test will be relevant and also that there exists the necessary scientific background. This last requirement may be a weak point in the whole performance testing approach, since our objective knowledge of in-use conditions is quite limited. Maybe more effort should be devoted to study activities and stresses before work on developing performance testing methods is accelerated. In this connection it should also be considered how feed-back from real buildings in a systematic way can be brought to use in the laboratories developing or revising Performance Test Methods.

COMPARISON BETWEEN SOME EXISTING PERFORMANCE REQUIREMENTS FOR AIR PERMEABILITY AND WATER-TIGHTNESS IN BUILDINGS

by ir R D'Havé and ir P Spehl
SECO, Brussels

SCOPE

This paper compares some existing performance requirements for the air permeability and watertightness of windows, and illustrates the necessity of further work towards harmonisation.

1 Air permeability of windows

The performance of air permeability is the flow of air blowing through the joints of the window.

Table 1 gives a list of the performance assessment methods, grading systems and requirements considered. The assessment methods are basically the same for all.

The grades (or classes) are drawn on Figures 1 and 2.

Both figures represent the flow Q depending on the static pressure difference Δp but on Figure 1, the flow Q is expressed in $m^3/h.m$ of joint, and in Figure 2 in $m^3/h.m^2$ of surface.

The figures define UEAtc classes A1, A2 and A3, and some other systems of classes have been added for comparison:

- The Danish Building Research Institute banded levels (defined at a pressure of 700 Pa) on both figures
- The British Interdepartmental Sub-Committee for Component Co-ordination grades on both figures.
- The French Standard NF P20–302 on Figure 1 (same as UEAtc classes)
- The Belgian STS 52 draft of 1979 on Figure 1 (UEAtc grade Al has been considered as not sufficiently severe and a subdivision of grade A2 has been considered necessary).
- The Dutch standard NEN 3661 on Figure 1
- The Norwegian Building Research Institute specifications on Figure 2
- The Israeli specifications on Figure 2

The UEAtc, the French, the Belgian and the Norwegian classes are defined according to the physical law $Q = m\Delta_p{}^{2/3}$ ie by the coefficient $m = \dfrac{Q}{\Delta_p{}^{2/3}}$ (assuming the the joint profiles do not change with variation in pressure). The others are classes of flow Q defined for a constant pressure Δ_p (DK, NL) or classes of pressure Δ_p defined for a constant flow Q (GB, IS).

Air permeability or tightness is not a user requirement as such: this performance is required from a window to satisfy the three following user requirements:

1 control of air temperature and velocity of the flow from the windows inside the rooms
2 heating energy saving
3 ventilation of the rooms (if no other installation provides it)

The third requirement is obviously in contradiction to the two first ones; therefore, if there is no ventilation installation, the final requirement for the air permeability of the window will need to be a compromise. The air permeability requirement for a closed window, being a function of the wind pressure, depends on the return period considered, the region (see the difference between Norway and Israel), the height above ground, the roughness of the terrain and the protection of the window from the wind by other parts of the same building or by other buildings.

The following example gives the proposed Belgian requirements for the new draft of STS 52 (1979):

a **non-opening parts:**
   $Q \leqslant 0.3\ m^2/h.m^2$ at $\Delta p = 100Pa$

b **opening parts** (windows, doors)

Height above ground	normal insulation (single glazing)	improved insulation (double glazing)
0 to 10 m	PA1	PA2
10 to 25 m	PA2	PA3
> 25 m and sea shore	PA3	PA3

2 Water-tightness of windows
The performance of water-tightness is expressed by the highest air pressure for which there is no infiltration of driving rain water in the room.

Table 2 gives the list of performance assessment methods, grading systems and requirements considered. Figure 3 shows the differences in the assessment methods: the Danish and Norwegian methods use pulsation of air and a small quantity of water representing driving rain, whereas the other five use static pressure and a larger quantity of driving rain.

Large differences are also to be noticed in the grading systems.

The following example gives the proposed Belgian requirements for the new draft for STS 52 (1979):

Height above ground	Protected facade (protrusions of 1.2 m)	Non-protected facade (no protrusion)
0 to 10 m	E2	E3
10 to 25 m	E3	E4
> 25 m and sea shore	E4	E5

For E4 and E5, there may be no water penetration inside the building under 250 pulsations of air from 0 to 250 Pa.

For curtain walls and h > 25 m, the test pressure is the pressure of the maximum normal wind according to NBN 460 ($\simeq$ wind obtained every year).

Table 1 Air permeability – windows – References

Country	Assessment method	Grades	Requirements
CEN	EN 42	–	–
UEAtc	Directives pour l'agrement des fenetres	Directives pour l'agrement des fenetres Grades: A1, A2, A3	–
B	NBN B 25–204 (EN42)	–	–
	STS 36 : tome I – 00.38.10 (metalwork)	–	STS 36 : tome III 36.10.12
	STS 52 : tome I 00.38.10 (woodwork) (draft 1979)	STS 52 : tome III 52.04.12 (draft 1979)	STS 52 : tome III 52.04.12 (draft 1979)
DK	YEB 2	YEB 2 banded levels	–
F	NF P20–501 (EN 42)	NF P20–302 Grades : A1, A2, A3	DTU 36.1/37.1
GB	ISCC technical note no 1	–	ISCC technical note no 1
	BS 5368 Part 1 (EN 42)	DD4	DD4
IS	Performance specifications for building elements	Performance specifications for building elements grades : types 1, 2, 3	Performance specifications for building elements
N	Vinduer av tre (NBI anvisning 10)	Vinduer av tre (NBI anvisning 10) grades 1, 2, 3, 4	–
NL	NEN 3661	NEN 3661 grades B15, B40, B100, K15, K40, K100	NEN 3661

Table 2 Water-tightness – windows

Country	Assessment method	Grades	Requirements
CEN	EN 86	–	–
UEAtc	Directives pour l'agrément des fenêtres	Directives pour l'agrément des fenêtres grades E1, E2 E3, E4	–
B	NBN B25–209 (EB 86)	–	–
	STS 36 – tome I 00.38.20 Menuiseries métal – liques	–	STS 36 tome III 36.10.12
	STS 52 – tome I 00.38.20 Menuiseries extérieures en bois (draft 1979)	STS 52 – tome III 52.04.12 grades PEau 1, PEau 2 (draft 1979)	STS 52 – tome III 52.04.12 (draft 1979)
DK	YEB 2	YEB 2 banded levels	–
F	NF P20–501 (EN 86)	NF P20–302 grades E1, E2, E3, E_E	DTU 36.1/37.1
GB	BS 5368 Part 2 (EN 86)	DD4	DD4
IS	Performance specifications for building elements	Performance specifications for building elements	Performance specifications for building elements
N	Vinduer av tre NBI anvisning 10	Vinduer av tre NBI anvisning 10	Vinduer av tre NBI anvisning 10
NL	NEN 3661	NEN 3661 grades B15, B40, B100, K15, K40, K100	NEN 3661

3 Conclusions

This study of some existing performance requirements for air permeability and water-tightness of windows shows that for both performances, a large variety of methods of test and grading systems is used.

This situation makes the comparison of the performance of products tested in different countries very difficult. Therefore a search for harmonisation is necessary. A first step has been the drafting of the European Standard for the methods of test for air permeability (EN42) and for water-tightness (EN86).

A further step should be taken to unify the grading systems, taking into account the large variety of climatic conditions existing throughout the world.

This could be a task for CEN, or even better for ISO.

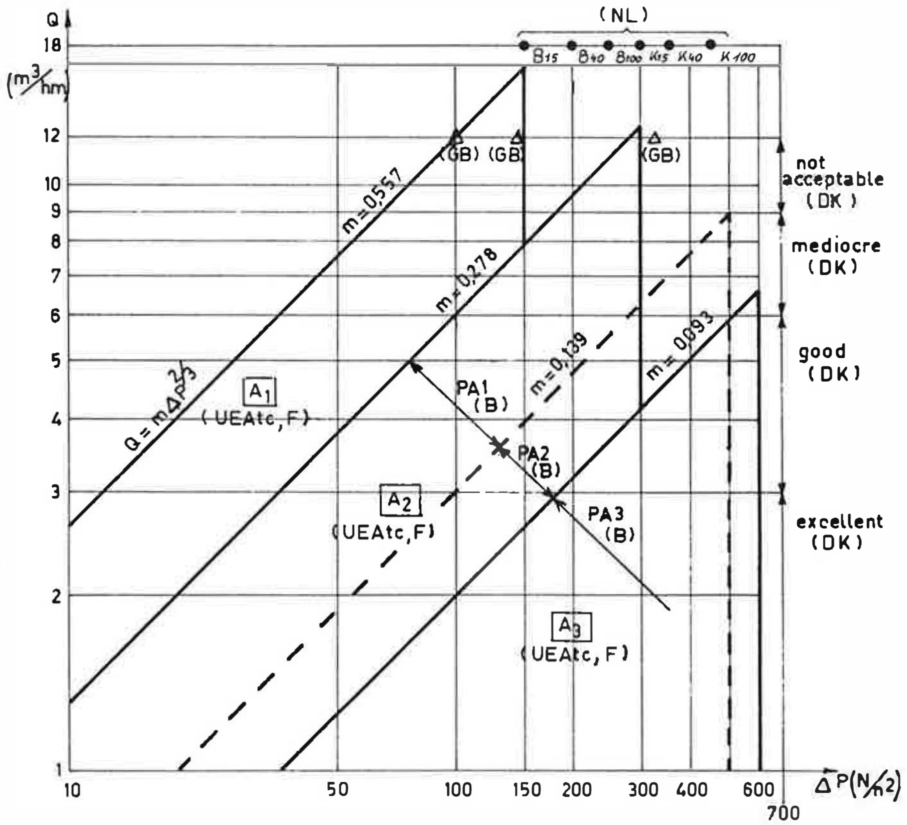


Figure 1

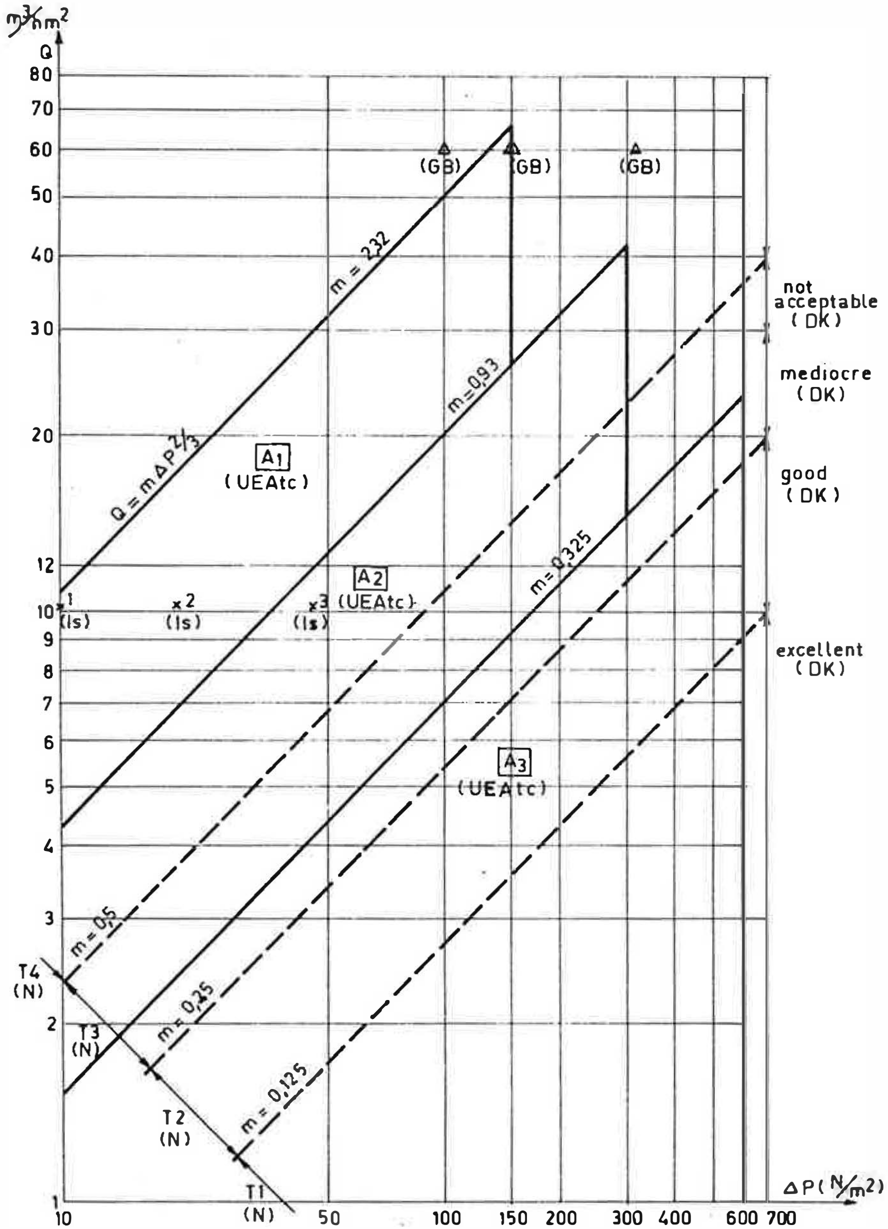


Figure 2

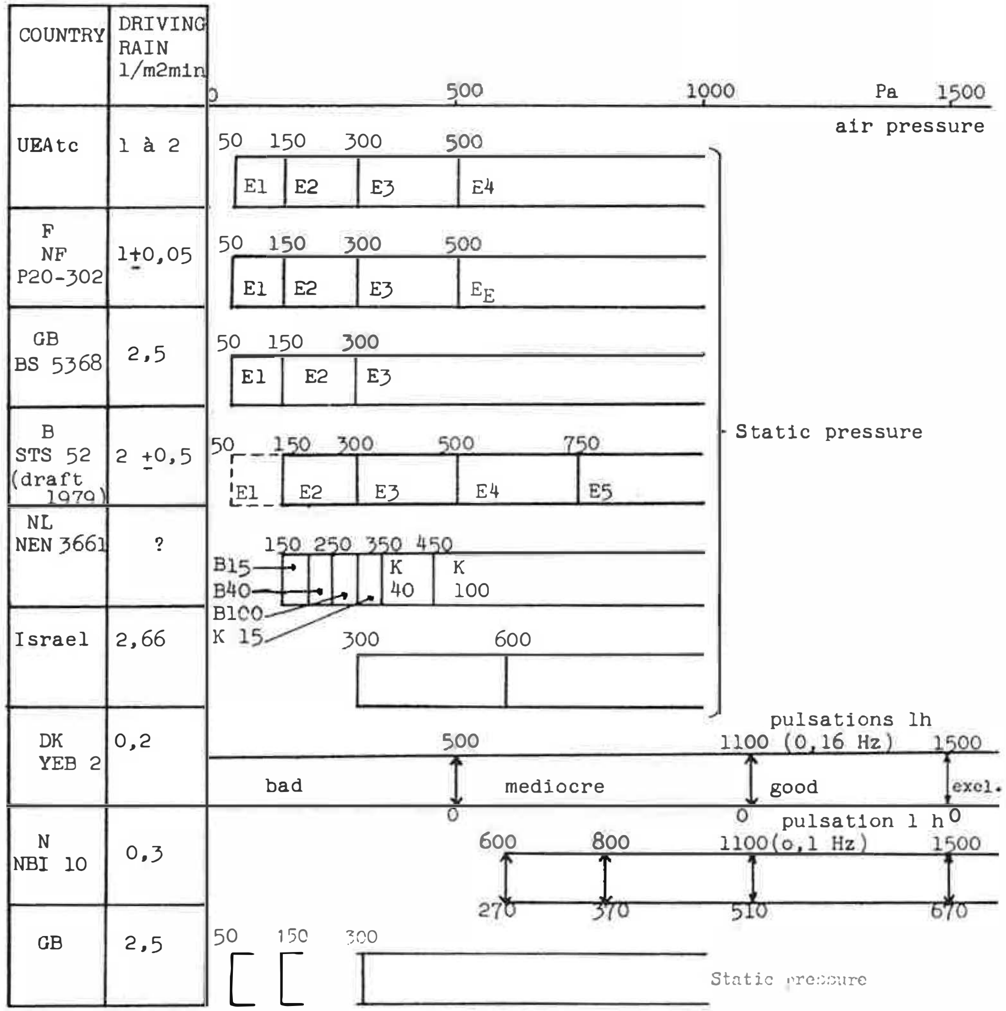


Figure 3 Water-tightness-windows

A NOTE ON THE SELECTION OF SAMPLES FOR PERFORMANCE TESTING
by J O May
Head of Materials Division
Agrement Board, UK

The application of the Performance Concept to buildings or parts of buildings invariably involves the necessity of testing for the various performance parameters required. The testing is not usually at the normal scale of laboratory-type testing, but involves large and expensive test facilities capable of taking a number of components mounted together as they would be in a building or a single component, such as a window, fixed into the same type of surround as it will meet in service. Such test facilities are rarely found in manufacturers' laboratories, but usually only exist in national test laboratories or in specialist commercial test houses. The limited availability of such test facilities, together with the high cost of such testing, makes it impracticable in most situations to consider carrying out sufficient tests to establish a statistical significance for the results.

There is also a more fundamental reason why it may be invalid to seek greater reproducibility and statistical interpretation in such large-scale performance testing. Such tests simulate some or all of the environmental conditions to which the product is expected to be subjected during its life, but each condition occurs on a probabilistic basis having a statistical distribution which may not be precisely known overall, and will certainly not be precisely known for any one given product installed in a given building. Examples are: maximum wind gust – once in fifty years? And what is it in the microclimate surrounding the building in question? – Or fire – what are the real-life statistics for the temperature/time curve, and what would it be for the building in question?

Performance testing is therefore carried out for the worst conditions that are reasonably expected to occur during the lifetime of the product when in use. The worst conditions are arrived at on the basis of judgement from such environmental data as exists. The test results may not appear as 'pass/fail' and may themselves again have to be interpreted against existing environmental data. These real uncertainties should, of course, be taken into account in design by the use of appropriate factors.

Statistically significant values for performance characteristics of products are therefore obtained only:

(a) where the performance test can be used as a quality control test for the product, eg, the measurement of air and water infiltration properties of windows in the UK. In this situation it is essential that the test method and apparatus is standardised, so that results between different manufacturers can be compared.

(b) where an easily measured physical property of the material can be used for quality control purposes and is directly related to a performance characteristic, eg, for a given material, the relationship between density and thermal characteristics.

It is doubtful whether, even if it were possible to obtain statistical data on performance characteristics, that the information in many instances would be useful. The variability of the product or system under laboratory test conditions may be of an entirely different order of magnitude to the variability in performance that will be achieved in buildings due to the vagaries of the building-in process and other factors such as building or component movements, ageing, etc.

The Agrement Board accepted at the start of its work on the assessment of building products that it could not attempt to obtain statistical data on performance characteristics except in very limited circumstances such as where design criteria have to be given for structural calculations. The following procedure has been adopted by the Board to minimise the risks involved in only testing one or at most a small number of units.

The essential prerequisite for the Board to undertake certification of a product or system is that it is under full-scale manufacture, that the product is under control, and that its use and installation can be defined.

For a product to be deemed under control, the following requirements are essential.

1 The raw materials are adequately specified, either in terms of the manufacturer's own acceptance specification, or by a grade of material manufactured to a fixed specification by a raw material supplier. The materials must be either checked by the manufacturer or supplied with a test certificate to show compliance with the specification. Records of tests must be kept.

2 The manufacturing process must be described in detail including formulations and methods of process control.

3 Quality control tests on the finished product must be carried out at sufficient frequency for the results to be statistically meaningful and the tests used must cover the important parameters as far as performance of the product is concerned. Records of the results must be kept and properly recorded.

4 The finished product must be specified in terms of upper and lower control limits on its important physical characteristics. The difference between the control limits must not be excessive.

5 All ancillary materials and components used with the product to install it on site must be adequately specified and defined.

Once it has been decided that the product and its installation has been adequately specified and that the product is under control, samples are selected for test. Before any testing is started, the test samples are checked against the manufacturer's own specification.

Tests are usually carried out in the following situations:

1 At the extremes of the product specification, eg, for a pipe joint, with sizes of socket, spigot and gasket for maximum and minimum interference.

2 Whenever changes in the product specification occur.

3 Whenever changes in the installation specification occur.

The above procedure is not perfect, but, in our opinion, it is as valid as any other procedure that does not involve a proper statistical approach to the selection of samples for test. The decision to select a number of samples at random from the manufacturer's stock does not give any more significant results unless these can be related to defined positions in the manufacturer's control charts, and they can be shown to cover an adequate range of the manufacturing variability.

PERFORMANCE TESTING AND THE USE OF STATISTICS

by Georg Christensen
Danish Building Research Institute

GENERAL

From a strictly scientific point of view it is not satisfactory to evaluate the results from Performance testing when only one test is carried out. Unfortunately this is often the case due to economic reasons. However, instead, a number of tests ought to be carried out in order to get a more realistic background for evaluation of a product.

It goes without saying that the use of a performance specification calls for development of prefabricated components with certain performances, and consequently commonly available methods of industry ought to apply for the evaluation of such building products. This means that quality testing procedures should be based on a statistical background, just as in other industries. The full scale, and hence once only, testing procedure should be considered merely as a kind of screening procedure.

It might now be considered that a statistical approach would need the development of a new theory within the Performance Concept. Fortunately this is not the case, since the traditional statistical methods used in general apply also when Performance test results are evaluated. This will be demonstrated in the following examples:

PERFORMANCE TESTING

Results from Performance testing can come from such different areas as measurement of air leakage of windows, indentation marks on surfaces, wear on floor coverings, and load-bearing and deformation characteristics of floor components. As an example, the statistical methods for strength and rigidity for point loads on a floor component are dealt with. In order to explain the procedure, a few definitions are needed, and these are now given.

DEFINITIONS

a Strength is given at the 'serviceability limit state' (cracking load P_c) and at the 'ultimate limit state' (ultimate load P_u).

b Rigidity is given by the deflection for a short-term point load of 1.0 kN.

The characteristic value of the strength P_{ck} and P_{uk} are chosen as the 5% fractile. This is the value below which not more than 5% of the strength values of a large quantity will be found. The characteristic value of the deflection u_k as the upper 30% fractile. This is the value above which not more than 30% of the deflection values of a large quantity will be found.

When only a small number of tests is carried out, the characteristic values should be determined at a significance level of not less than 75%. The significance level denotes the minimum probability that a quantity is rejected if it does not meet the requirements.

STATISTICAL EVALUATION

Whenever possible, test results should be evaluated based on a number of test results which are treated statistically. It should here be noted that the test results should be 'normally distributed'. If this is not the case, other statistical methods should be investigated.

Assuming a normal distribution of the test results, the characteristic values can be calculated according to:

$P_p = (1 - k_p\,\sigma)\,\bar{P}$ and

$u_p = (1 + k_p\,\sigma)\,\bar{u}.$

$\bar{P}$ and $\bar{u}$ are the simple mean values, σ is the variation coefficient, p = 5 for the 5% fractile, and k_p is a coefficient which is dependent of the number of test results according to the following table:

k_p	Number of tests								
	4	5	6	8	10	15	20	40	100
k_5	2.68	2.46	2.33	2.18	2.10	1.99	1.93	1.83	1.75
k_{30}	1.06	0.96	0.91	0.84	0.79	0.74	0.70	0.65	0.60

If the testing gives extremely high or low values for the variation coefficient, the reason for such abnormalities should be investigated. Such an investigation might lead to a certain correction of σ before the calculation of the characteristic value is made.

EXAMPLE
A subfloor made of 22 mm chipboard is tested in order to evaluate its resistance to point loads and the deflections for similar loads.

The chipboard is placed as the top layer of a component with rafters underneath. The testing is carried out on the free span between two rafters, and it is assumed that six tests can be carried out on the same component without having the test results interfering with one another.

The ultimate loads were recorded as follows:

P_u : 3.34 – 3.60 – 3.58 – 3.59 – 3.35 – 4.06 kN

The cracking loads were recorded as follows:

P_c : 3.05 – 3.28 – 3.32 – 3.08 – 3.15 – 3.56

The deformations measured for a certain load 1.0 kN were the following:

u = 2.37 – 2.50 – 2.26 – 2.23 – 2.23 – 2.46 mm.

Calculations of test results will be the following:

a Characteristic ultimate load P_{uk}

$$\bar{P} = \frac{3.34 + 3.60 + 3.58 + 3.59 + 3.35 + 4.06}{6} = 3.59 \text{ kN}$$

$$s = \sqrt{\frac{\Sigma\,(P - \overline{P})^2}{5}} = 0.26 \text{ kN}$$

$$\sigma = \frac{s}{\overline{P}} = \frac{0.26}{3.59} = 0.073$$

$$P_{uk} = 3.59\ (1 - 2.33 \cdot 0.073) = 2.98 \text{ kN}$$

b Characteristic cracking load P_{ck}

$$\overline{P} = \frac{3.05 + 3.28 + 3.32 + 3.08 + 3.15 + 3.15 + 3.56}{6} = 3.24 \text{ kN}$$

$$s = \sqrt{\frac{\Sigma\,(P - \overline{P})^2}{5}} = \sqrt{\frac{0.19}{5}} = 0.19$$

$$\sigma = \frac{s}{\overline{P}} = \frac{0.19}{3.24} = 0.059$$

$$P_{ck} = (1 - k_5\ \sigma)\,\overline{P}$$

$$P_{ck} = (1 - 2.33 \cdot 0.059)\ 3.24 = 2.80 \text{ kN}$$

c Characteristic value of deflection

$$\overline{u} = \frac{2.37 + 2.50 + 2.26 + 2.23 + 2.23 + 2.46}{6} = 2.34 \text{ mm}$$

$$s = \sqrt{\frac{\Sigma\,(u - \overline{u})^2}{5}} = 0.12 \text{ mm}$$

$$\sigma = \frac{s}{\overline{u}} = \frac{0.12}{2.34} = 0.051$$

$$u_{30} = 2.34\ (1 + 0.91 \cdot 0.051) = 2.45 \text{ mm}$$

CONCLUSIONS IN PLAIN LANGUAGE

a P_{uk}

When taking a group of components eg 100, there will be a 75% probability that 5 components or less will have an ultimate strength of 2.98 kN or less.

b P_{ck}

When taking a group of components eg 100, there will be a 75% probability that 5 components or less will have a cracking strength of 2.80 kN or less.

c u_k

When taking a group of components eg 100, there will be a 75% probability that 30 components will have a deformation of 2.45 mm or more for a load of 1.0 kN.

THE VARIABILITY OF TEST RESULTS WHEN ASSESSING THE RESISTANCE OF WINDOWS TO WATER AND AIR PENETRATION USING BS4315

by J F S Carruthers and C J Newman
Princes Risborough Laboratory of the
Building Research Establishment, UK

INTRODUCTION

The investigation was undertaken to assess the variability of windows tested to British Standard 4315:Part 1:1968, Methods of test for resistance to air and water penetration: windows and gasket glazing systems. Since the issue of this Standard, it has been widely used as a means of specifying the performance of windows for use in the UK. However, the variability of this test method had not been examined systematically.

The programme was intended to assess variability in both water and air tests from five causes:

Different designs of window
Differences between individual windows of the same type
Different pressure test boxes
Different test operators
The residual variability inherent in the test after excluding all other factors

THE WINDOWS TESTED

One hundred and thirty windows, approximately 1.2 m square and of the five types described in Table 1, were tested.

THE PRESSURE TEST BOXES

Three different pressure boxes were used, though the majority of the tests were carried out on two. Two skilled operators were involved.

RESULTS

The main results for air penetration are shown in Tables 2 and 3 and for water penetration in Tables 4 and 5.

CONCLUSIONS

The test programme has shown that considerable variation occurs in the testing of the air and water penetration attributes of windows to BS 4315:Part 1.

With both water and air penetration there was no single source of the wide variation found in the test results. Indeed, the differences between individual windows of a type, between boxes, between operators and associated with the test procedure as such appeared to contribute to a similar extent to the total variation experienced.

THE IMPLICATIONS OF THE TEST PROGRAMME FOR THE SPECIFICATION OF AIR AND WATER PENETRATION

With the range in variation found in the test programme, it is not possible to predetermine without an extensive background of test results the variation that could occur with a particular design and type of window. As a result, it becomes difficult to set realistic performance levels for specification and procurement purposes. One approach is simply to assume a maximum variation and apply it to all windows in setting performance levels. This approach, however, could penalise windows giving a consistent performance, especially in the case of water penetration. A second alternative is to adopt a two-stage statistical procedure that can accommodate the variation associated with the actual group of windows tested.

Table 1 Details of windows examined

Type		Material and finish	Draught stripping	Number obtained for testing
Designation	Description			
IA	Horizontal slider in timber sub-frame One opening light one fixed light	Aluminium mill finish	Brush and neo- prene seals	30
IB	Horizontal slider in timber sub-frame One opening light one fixed light	Aluminium mill finish	Brush seal	5
II	Side-hung casement in timber sub-frame One opening light one fixed light Weather bar at head	Aluminium mill finish	Neoprene seal	40
III	Vertical slider in timber sub-frame Two sliding lights	Aluminium mill finish	Brush and neo- prene foam seal	20
IV	Side-hung casement and vent light Two opening lights one fixed light Weather bar at head	Timber Gloss paint finish	None	30
V	Side-hung casement and vent light in timber sub-frame Two opening lights one fixed light Weather bar at head	Steel Gloss paint	None	6

Table 2 Air penetration at 100 Pa of windows tested using two boxes and two operators

Type designation	Number of windows tested	Mean air penetration (m^3/hm)*	Standard deviation ($m^3/h\,m$)*	Range within which 95% of window tests occur ($m^3/h\,m$)*
IA	30	3.6	1.6	0.4 – 6.8
IB	5	13.8	1.8	10.2 – 17.4
II	40	2.2	1.3	0 – 4.8
III	20	11.1	1.6	7.9 – 14.3
IV (a) whole window	30	7.9	3.2	1.5 – 14.3
(b) side-hung casement only	30	8.7	4.2	0.3 – 17.1
(c) vent light only	30	6.4	3.1	0.2 – 12.6
V (a) whole window	6	23.8	3.8	16.2 – 31.4
(b) side-hung casement only	6	25.0	4.5	16.0 – 34.0
(c) vent light only	6	21.8	4.1	13.6 – 30.0

*The rate of air penetration is expressed at m^3/h per metre length of opening light

Table 3 Source and extent of variation in the air penetration of windows

Type designation	Source and extent of variation: Between windows: Number of windows examined	Between windows: Standard deviation ($m^3/h\,m$)	Between two box/operator combinations: Number of windows examined	Between two box/operator combinations: Difference in mean value ($m^3/h\,m$)	Associated with test procedure: Number of windows examined	Associated with test procedure: Standard deviation ($m^3/h\,m$)
IA	20	0.9	10	1.3	20	1.3
IB	5	1.3	–	–	5	0.8
II	20	1.3	10	1.3	20	0.8
III	20	0.9	10	1.4	20	1.1
IV (a) whole window	20	2.1	10	1.3	20	1.3
(b) side-hung casement only	10	3.5	10	3.3	10	0.8
(c) vent light only	10	2.7	10	2.0	10	0.9
V (a) whole window	6	*	–	–	6	2.5
(b) side-hung casement only	6	*	–	–	6	3.7
(c) vent light only	6					

*No statistically significant difference at the 5% level
–No test carried out

Table 4 Water penetration (pressure at which 'gross' leakage occurred) tests on windows using two boxes and two operators

Type designation	Number of windows tested	Mean pressure for 'gross' leakage (Pa)	Standard deviation (Pa)	Range within which 95% of window tests occur (Pa)
IA	30	160	55	50 – 270
IB	5	190	22	146 – 234
II	40	360	241	0 – 842
III	20	630	184	262 – 998
IV (a) whole window	30	200	102	0 – 404
(b) side-hung casement only	30	230	101	28 – 432
(c) vent light only	30	310	259	0 – 828
V (a) whole window	5	260	22	216 – 304
(b) side-hung casement only	5	280	27	246 – 314
(c) vent light only	5	260	22	216 – 304

Table 5 Source and extent of variation in water penetration of windows

Type designation	Source and extent of variation							
	Between windows		Between two operators		Between two boxes		Associated with test procedure	
	Number of windows	Standard deviation (Pa)	Number of windows	Difference in mean value (Pa)	Number of windows	Difference in mean value (Pa)	Number of windows	Standard deviation (Pa)
IA	6	10	10	*(1)	10	*(1)	6	29
IB	5	19	–	–	–	–	5	25
II	5	253	10	*(1)	10	*(1)	5	209
III	5	70	8	63	8	144	5	125
IV (a) whole window	5	*	5	80	5	*	5	97
(b) side-hung casement only	5	*	5	130	5	40	5	50
(c) vent light only	5	89	5	*	5	95	5	114
V (a) whole window	5	23	–		–		5	24
(b) side-hung casement only	5	22					5	20
(c) vent light only	5	23					5	24

*No statistically significant difference at the 5% level
–No test carried out

(1) For window types IA and II differences between operators and between boxes were not isolated. The result given for the window types refers to the difference between two box/operator combinations.

SUGGESTED TEST PROCEDURES

It is suggested that the criteria for acceptance should be such that there is a 75 per cent chance that the group of windows tested is drawn from a population of which 95 per cent of the windows have a higher test result than the specified level. It can be shown[2] that this criterion is met when the mean of five results is more than 2.46 times the standard deviation of those 5 results above the specified level. With fewer test samples the multiplying factor is increased, for example for four windows it is 2.68, and for more test samples the multiplying factor is decreased, for example for 6 windows it is 2.33. The use of five samples is suggested in this paper as being appropriate taking into account, on the one hand, the need to minimise the cost incurred by testing and on the other, the benefits obtained from testing a larger number of windows.

If the five windows tested fail this initial coarse assessment, it is permissible to examine another five windows. The tests on these ten windows are then repeated to remove the effect of the variability of the test procedure and the mean result when reduced by 2.1 standard deviations should then be above the specified level for a 'pass' to be accepted.

The proposed method is detailed and an example of the calculation for the water penetration attribute is given in the Appendix to this paper.

REFERENCES

1 The repeatability and reproduceability of test results on windows and wall span elements and the expected results. By J F S Carruthers and C J Newman. BRE Current Paper CP 49/77.

2 A note on performance testing and use of statistics. By G Christensen, Danish Building Research Institute. Paper 10/10, CIB W60, Oslo, March 1977.

APPENDIX

SUGGESTED ASSESSMENT PROCEDURE FOR AIR AND WATER PENETRATION TESTS

Stage 1 Calculation of characteristic test performance from results of testing five windows once each.

Let $x_1\ x_2\ \ldots\ldots\ x_5$ be individual results

the mean $\bar{x} = \dfrac{\Sigma x}{5}$

and standard deviation $s = \sqrt{\dfrac{\Sigma(x^2) - \dfrac{(\Sigma x)^2}{5}}{4}}$

With a 75 per cent chance that this group of windows is drawn from a population of which 95 per cent of the windows have a higher or lower test result than the specified level then:

the characteristic test performance = $\bar{x} \pm 2.46s$

For example if five windows gave the following results for the water penetration test:

150 200 150 200 250 Pa

$\Sigma x = 950, \bar{x} = 190$ and $\Sigma(x^2) = 187500$

$$s = \sqrt{\frac{187500 - \dfrac{950^2}{5}}{4}} = 41.8$$

Therefore characteristic water test performance = 190 – (2.46 x 41.8) Pa
= 87 Pa

If the specification pass level is 50 Pa the windows would pass. However if it was 100 Pa they would just fail and stage 2 of the assessment procedure could be invoked especially as the characteristic water test performance approaches the pass level required.

Stage 2 Calculation of characteristic test performance from results of testing ten windows twice each:

Let $x_{1.1}\ x_{1.2}\ \ldots\ldots\ldots\ x_{1.10}$ be first test results

and let $x_{2.1}\ x_{2.2}\ \ldots\ldots\ldots\ x_{2.10}$ be second test results

Then to use the analysis of variance techniques to identify the window variance:

(a) square each test result and sum, ie $\Sigma(x^2)$

(b) add together two test results for each window, square these totals, sum the squares and divide by 2, ie

$$\frac{(x_{1.1} + x_{2.1})^2 + (x_{1.2} + x_{2.2})^2 + \cdots (x_{1.10} + x_{2.10})^2}{2}$$

(c) add together all twenty test results, square this total and divide by 20 ie

$$\frac{(\Sigma x)^2}{20}$$

Then test variance $\quad s_t^2 = \dfrac{(a) - (b)}{10}$

As the variance between results for different windows

$$2s_w^2 + s_t^2 = \frac{(b) - (c)}{9} \text{ where } s_w \text{ is the window variance}$$

$$\text{then window variance alone } s_w^2 = \frac{\frac{(b) - (c)}{9} - \frac{(a) - (b)}{10}}{2}$$

$$\text{and characteristic performance now} = \bar{x} \pm 2.1\, s_w$$

For example if two water penetration tests on each of ten windows give the following results

1st test	150	200	150	200	250	150	150	200	150	200 Pa
2nd test	150	150	200	200	200	200	100	150	150	100 Pa
then $(x_1 + x_2)$	300	350	350	400	450	350	250	350	300	300 and $\bar{x}$ = 170 Pa

(a) $\Sigma(x^2) = 605\ 000$

(b) $\Sigma\frac{(x_1 + x_2)^2}{2} = 592\ 500$

(c) $\frac{(\Sigma x)^2}{20} = 578\ 000$

$$\text{Therefore test variance } s_t^2 = \frac{605\ 000 - 592\ 500}{10} = 1250$$

As variance between results for different windows

$$2s_w^2 + s_t^2 = \frac{592\ 500 - .57\ 800}{9} = 1611.1$$

$$\text{Then window variance alone } s_w^2 = \frac{1611.1 - 1250}{2} = 180$$

$$\text{and } s_w = 13.4$$

$$\text{Therefore characteristic performance} = 170 - (2.1 \times 13.4)$$
$$= 142 \text{ Pa}$$

The windows now pass the 100 Pa specified level.

EXPRESSING PERFORMANCE VALUES IN BANDED LEVELS

by Klaus Blach
Danish Building Research Institute

The performance documents considered here are primarily 1) general lists of functional requirements as published for example by research institutions, 2) performance specifications worked out by consultants and designers for actual projects and 3) product information sheets – containing performance statements – as issued by firms and building centres.

In all three types of performance documents it is necessary somehow to state the levels or values of performance which are respectively presented, requested or offered.

It probably needs no explanation that collaboration will be facilitated and misunderstandings can be avoided if the parties involved can indicate performance levels in a common 'language'.

STATING SINGLE VALUES IS SELDOM SATISFACTORY

If a statement like 'the partition should be ½–brick' is substituted for by functional requirements, as, for example, regarding strength, fire resistance and sound insulation, the performance concept approach has of course been applied in principle. Experience has shown, however, that it is seldom satisfactory to state only a single value for each functional requirement, firstly because performance statements thereby tend to become as normative as the old descriptive statement, and secondly because a single-value-statement as a rule will correspond only poorly with the complex conditions met with in daily practice.

As just one example, the air- and rain-tightness for a window under a large overhang in a one-storey house does not usually have to be as good as that for a much more exposed window on the tenth floor in a high building.

Other conditions which make the single-value-statement obsolete or incorrect in this case may be, for example, regional differences as to precipitation and wind velocities.

BANDED LEVELS OF PERFORMANCE ARE BETTER

For institutions which publish general lists of functional requirements, it is a 'must' to work with banded levels of performance, quite simply because the documents in question should be applicable to a multitude of different – and unknown – projects. Consultants and designers working on specifications may sometimes know quite definitely which specific performance-value they want or need for a given project. But a trade-off between cost and performance may often make it desirable also for such users to have several performance-values as possibilities.

Also the firms (manufacturers, contractors etc) will be able to use banded levels of performance to their advantage. They can thus present solutions at various levels of performance – and at various levels of cost.

The most important fact in favour of the banded levels of performance, is that their use will leave the final choice of performance value with the only persons qualified to make the choice: those who know the actual project in hand.

FUNCTIONAL REQUIREMENTS FOR BANDED LEVELS

There has so far been no agreement as to a standard expression of performance values in banded levels. One reason is possibly that there are quite a few requirements which such banded levels should satisfy – and also some practical difficulties to be surmounted.

In the following are discussed a series of functional requirements for such banded levels and some practical solutions are suggested.

PERFORMANCE LEVELS MUST CORRESPOND TO ACCURACY OF EVALUATION METHOD

Accurate limit values may be found in building regulations and codes, but are seldom applicable in connection with evaluation based on testing, calculation or subjective judgement.

Even the best performance testing method will as a rule not allow an evaluation which is 'accurate to the fourth decimal' or which makes it possible to draw thin, accurate borderlines between the quality – or performance-categories.

Banded levels of performance should mirror the above-mentioned facts of life, for example by not being sub-divided too finely, by figures being rounded-off as necessary, and by masking of borderlines between categories.

OPEN-ENDED SCALE – TRUNCATED FOR USE IN PRACTICE

The scales containing the banded levels should as a matter of principle be open-ended, because as long as they are not applied to specific projects it is in many cases not possible to indicate the appropriateness of specific levels or values. At the equator thermal insulation – to prohibit heat escaping from the house – can thus be nil, while under perma-frost conditions human survival requirements may necessitate extreme thermal insulation to be applied regardless of cost.

In most countries or regions it will, however, as a rule be possible to cut out of the open-ended band a segment which contains a sufficient number of levels or values.

At the 'bottom' of the band should be found the level or value which indicates a quality below the acceptable, and at the other end of the band should be found the level or value beyond which it is uneconomic to increase quality.

In countries which encompass considerable climatic variations, either because of their size or their topography, segments regarding, for example, precipitation or wind may have to be rather wide so as to encompass a sufficient number of levels. On the other hand there are probably other segments which could find universal use without being especially wide. As an example, it seems as if certain space requirements and also some requirements as to dimensional compatibility are much the same all over the world.

VALUE-LOADED DESIGNATIONS OF LEVELS MUST BE AVOIDED

As explained above, it may be the case that a 'barely acceptable' value for the air-or rain-tightness of a window is quite 'good enough' if the window is positioned under a large overhang in a one-storey building: and vice versa at the other end of a band of levels a very high performance value may be necessary to assure a satisfactory result.

When choosing how to designate levels, it must be considered that it would hardly be acceptable to, for example, manufacturers if they were asked to market any solution for which a level of performance would have to be stated as 'low' or 'barely acceptable' – even in the cases where it could be argued that such a performance value would be perhaps more than good enough.

This would indicate that a completely neutral set of designations for levels would be appropriate, as for example 5, 4, 3, 2, 1; or K, L, M, N, O. Such designations are, however, not always perfectly satisfactory, because they may necessitate explanations in order to be understood correctly.

As a suggested solution, it is proposed to use slightly different but coordinated designations in connection with the three types of performance document considered here.

A DESIGNATION OF LEVEL MUST BE EASY TO PRINT, READ AND UNDERSTAND
A level can be designated using a phrase, a word, a letter, a figure or a symbol.

Phrases and words immediately can be understandable, but as mentioned above, they tend to become value-loaded, and they take up relatively much space. This is undesirable, especially in performance specifications and product information publications.

Designation through letters and figures will be an advantage in printing and also easy to read, but understanding may require further (time-and space-consuming) explanations.

Symbols can be easy to read and understand, but may have shortcomings when the 'printing' is based on typing, as it would seem to be difficult to find a usable set of symbols on normal typewriter key-boards.

A suggested solution is to use words, letters/figures and symbols in various combinations on the three types of performance document considered here. A special case can be argued for the use of symbols in product information publications, because they may facilitate the comparison of a long series of similar products which have to be presented in a very confined space.

The attached proposal for Banded Performance Levels is based upon the above considerations

---

DIFFICULTIES IN USING BANDED LEVELS
Scales to be stated on bands may be linear or logarithmic or arbitrary. It is suggested that this difficulty should not lead to use of bands subdivided in many different ways. For practical purposes it will be easier to work with series which are subdivided in a standard way.

Sometimes the performance in question will be best at a high value, in other cases at a low value. For ease of reading and understanding, it is suggested that bands should always read from 'un-acceptable' at the left towards 'excellent' at the right.

In a few cases a band would theoretically have to read from 'un-acceptable' over 'excellent' and to 'un-acceptable' again. As an example, this would occur with statements of appropriate working distances between tabletops in kitchens. Distances of 1.0 – 1.2 m allow for proper work spaces in two-sided kitchens, but distances both below and above these figures are less acceptable.

If banded levels are to be used in international performance documents, ie standards, it will probably be necessary to suggest how the open-ended bands may be extended in one or the other direction, in order to correspond with local conditions. Thus, a heat-insulation value which is considered excellent for two of the climate zones in France may be deemed only good in Denmark and barely acceptable in northern Norway. This again underlines the importance of avoiding value-loaded designation of performance levels.

Any presentation relying on the use of banded levels of performance could tempt its users – consultants or designers – to specify 'high levels' generally, just to be on the safe side. Any proposed wider use of banded levels should, therefore, be followed up by explanations that requirements for unnecessarily high performance may mean money squandered.

APPENDIX

PERFORMANCE PROFILES – A FUTURE POSSIBILITY

This Appendix presents an idea for using performance data expressed in banded levels in comparing various aspects of performance of competitive products. It is not regarded as an essential adjunct to the use of banded levels.

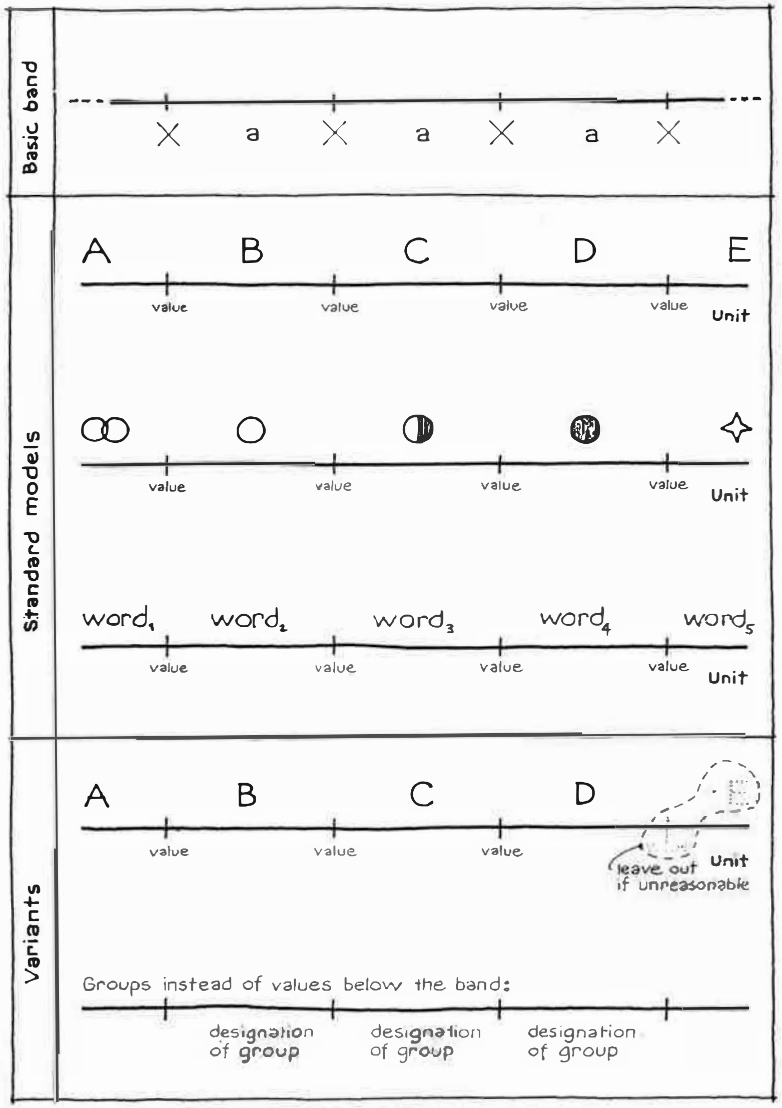
By a performance profile is here meant the visualisation of desired quality in a project which may be achieved by listing a series of banded levels with an indication of the value chosen for each item.

Work with such profiles was attempted previously in connection with the Break-Through Project in the USA. Further development of the idea will, however, have to rely upon agreement concerning a suitable band-design which can find wider use.

Performance profiles may be of special interest in daily practice, because they can facilitate a quick comparison of the content of a performance specification (desired performance profile) with offers received from various firms.

Note: The following diagrams employ the letters A, B, C, D and E to designate bands. They were prepared before W60's decision to recommend the use of the more neutral letter series J, K, L, M, N, O and P, noted in the text above and in the Introduction.

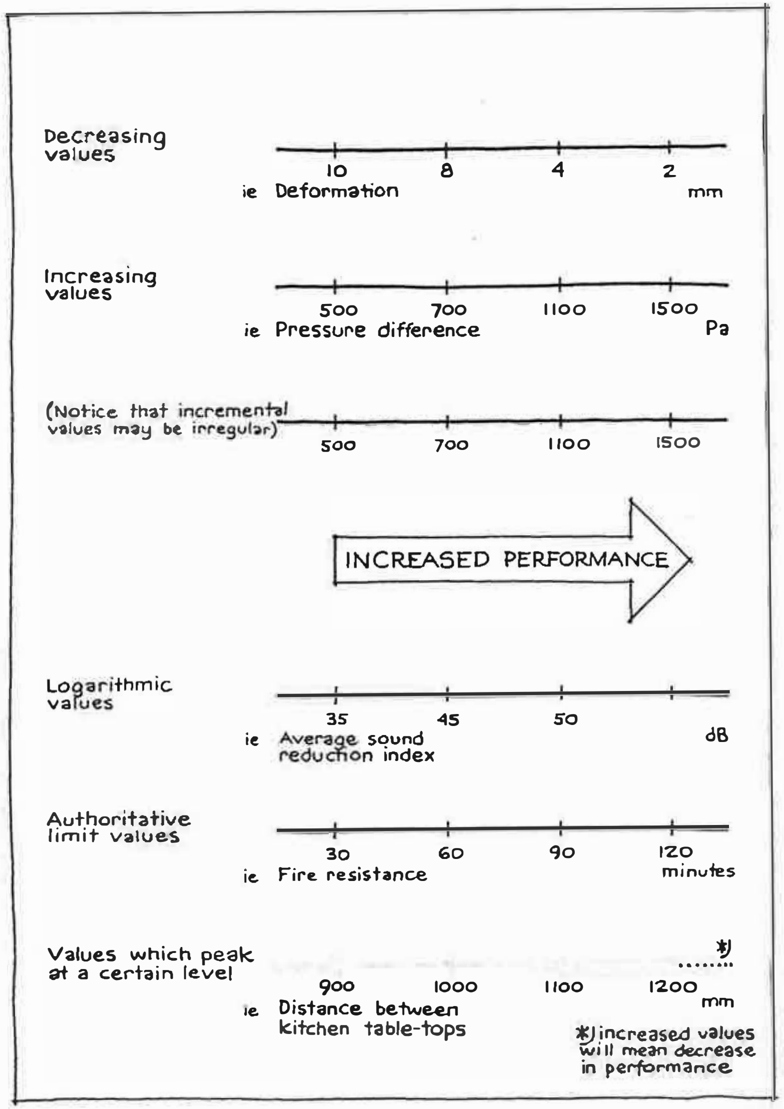
PROPOSED BAND DESIGNS.



ꝏ O ◑ ● ♠ ✧ as used in some consumers' journals and reports: ꝏ very bad; O mediocre; ◑ good; ● very good; ✧ excellent.

SBI	1
B-9/CIB W60 KB	78-02-11

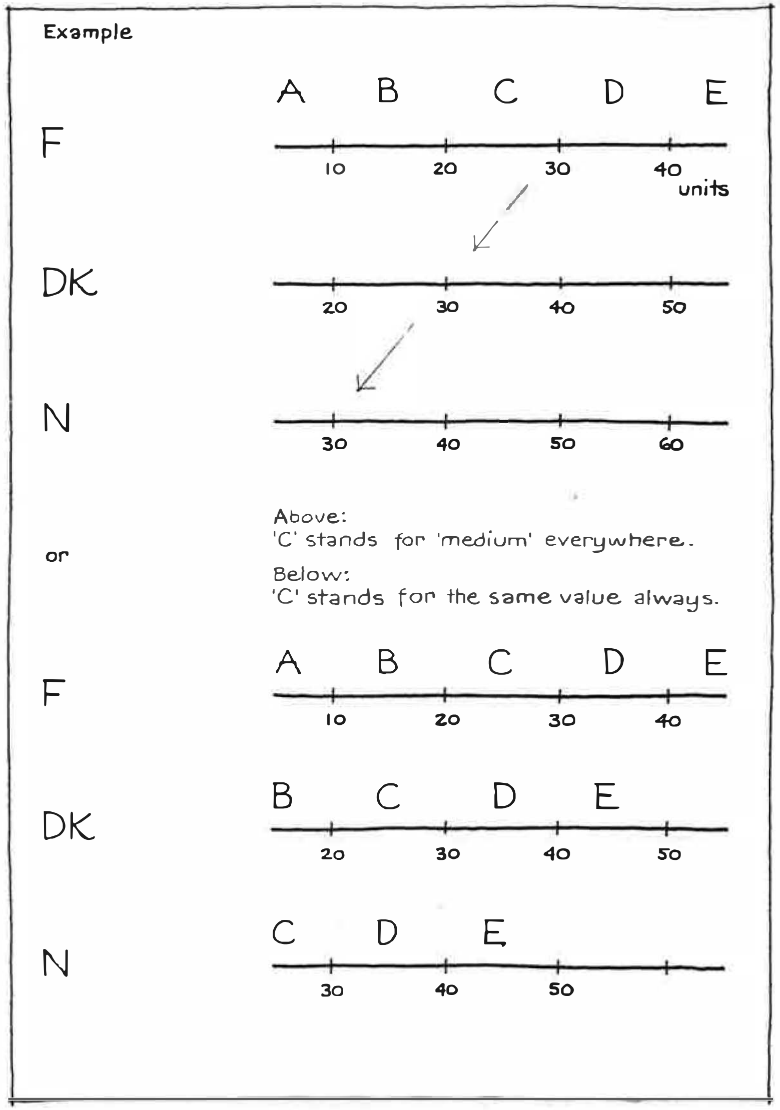
STANDARD BAND DESIGN FOR VARYING SETS OF VALUES



The use of standard band designs will make it easy to read banded levels of performance -- at a glance.

SBI	2
B-9/CIB W60 KB	78-02-05

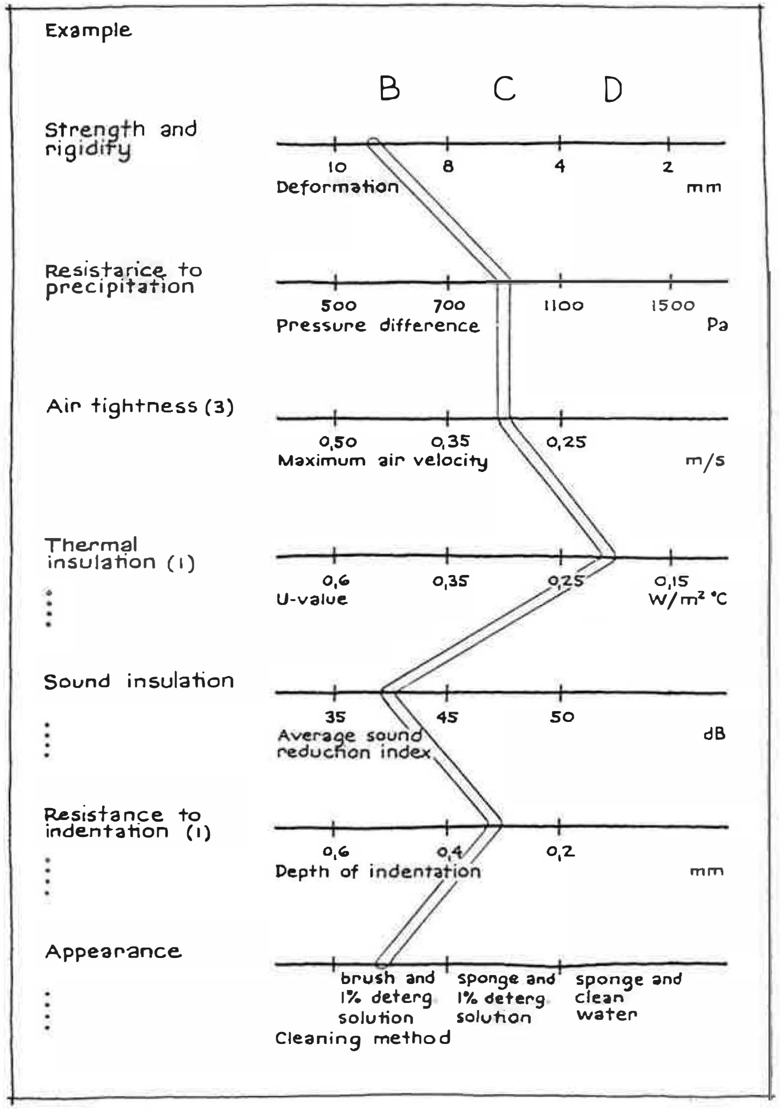
ADAPTATION OF BANDS



An international standard band design must in principle 'span from the poles to the equator'

SBI	3
B-9/CIB W60 KB 78-02-05	

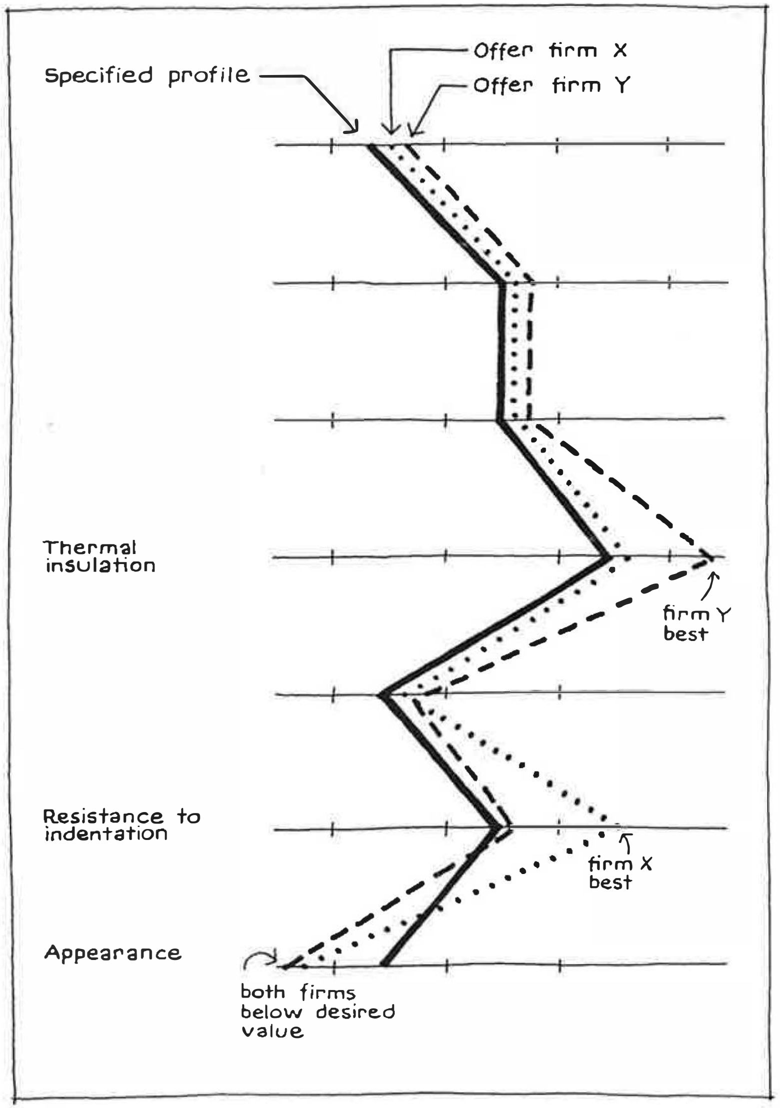
PERFORMANCE PROFILE



⋮ indicate that one or more properties have been left out

SBI	4
B-9/CIB W60 KB 78-02-05	

COMPARISON OF PERFORMANCE PROFILES



The designer must decide (1) is eventual extra cost of higher quality justified? (2) is thermal insulation more important than resistance to indentation? (3) can appearance-level be lowered?

SBI 5
B-9/CIB W60 KB 78-02-05