A THEORY OF VALIDATION FOR DERIVATIVE SPECIFICATIONS AND CHECK LISTS[1]

By Lee Byrne

Visiting Professor of Secondary Education, New York University

Part I. Research Products Which May Be Classified as Derivative Specifications and Check Lists

Meaning of Specification

In specification something is assigned a specific character. The something to be thus assigned a specific character may be called the specificandum. The specific character assigned to the specificandum, or (as a second meaning) the act of so doing, may be called the specification.

A proposition is the smallest unit in which it is possible to embody a complete thought and is ordinarily represented by a single sentence. In specification the characterization may be confined to a single proposition or it may be extended to include an indefinitely large number of propositions. So a specification may be embodied in a sentence, a paragraph, a chapter, or a whole book. No matter how far it is extended it will never give complete determination, as our knowledge cannot be made exhaustive or our control be given an absolute precision.

In view of the meaning assigned to specification it is evident that very many books and monographs could in this sense be classified as specifications.

Meaning of Derivative Specification

There is a type of specification (book or monograph) which is developed by deriving it from a group or class of specifications which already exist. This class may be a total class of all such specifications, or a group of those accepted as authoritative, or a group of those taken to be representative. A specification derived in this manner may be called a derivative specification. As an example we could take almost any first-class work by a present-day historian; by historians it would be called "secondary" because it is based on study of pre-existent documents called "primary sources."

Meaning of Check List

The act of *deriving* a product from a pre-existent set of documents may, as we have seen, take the form of a derivative specification, embracing an as-

[1] This paper is an amplification of a report made in the statistical section of the American Educational Research Association at its meeting in February, 1931.

semblage of determinates or determinations. On the other hand the product derived may be intended merely to indicate the ground covered or to be covered by determination, without actually selecting the particular determinations. Such a product will be called a check list. The term is not a very happy one, but it is in very common use. If we think of a specification as an assemblage of determinations then a check list could be thought of as a corresponding set of determinables.[2] Since any determinable is capable of an indefinite number of determinations it is evident that a long check list could give rise to an extremely large number of different specifications, of which, of course, some fraction might prove undesirable, inadmissible, or false.

Modes of Specification: How We Specify

If we examine any specification to see how the specifying is done we shall find that it ultimately takes the form of specification under aspects. The following diagram indicates the principal (perhaps all the) possibilities in the way of specification.

Naming the original or main specificandum
  Naming an aspect
    Characterization of the specificandum under the aspect named
    .
    .
    .
  Naming a relation (includes process, operation etc.)
    Naming an aspect of the relation
      Characterization of the relation under aspect named
    .
    .
    .
    Naming a relatum or thing related (a new specificandum)
      Naming an aspect of the relatum
        Characterization of relatum under aspect named
    .
    .
    .
  Naming a part
    Naming an aspect of the part
      Characterization of the part under aspect named
  .
  .
  .

(The naming of aspects may be merely implicit but it is always present in principle.)

[2] On the notion of the "determinable," which is due to W. E. Johnson, see his Logic, Cambridge University Press (1921), Part I, p. xxxv and Chapter XI.

Thus it appears that if specification is pressed far enough it always ultimately becomes specification under aspects. Aspect and determinable may be regarded as synonyms.

Current Examples of Derivative Specifications and Check Lists

At the present time it will be found that we have very many products of research which take forms capable of being classified as some kind of derivative specification or (derivative) check list in the senses in which these expressions have been explained.

I have distinguished more than twenty different logical types of derivative specification or check list which are exemplified in the current literature of educational research and related subjects. However space will not permit exhibition of examples of these different types.

Part II. Validation of Derivative Specifications and Check Lists

Many research products may be classified as derivative specifications or check lists, derivative in the sense that they have been derived from a group of documents (books, articles, journals, newspapers, courses of study, etc.) through analysis of their content. Such source documents themselves we shall call specifications or groups of specifications.

The only validation problem raised here is the question whether the resulting check list or derivative specification truly represents the class of source specifications used. The further question whether the class of source specifications itself constitutes a satisfactory source is not discussed.

From this point of view, if a check list or derivative specification is based in some suitable manner on *all* the documents of the class represented, no real validation problem arises; the validity has to be regarded as perfect.

It may often happen that the investigator does not wish to analyse *all* of the specifications of the class in question but prefers to save time and labor by confining his analysis to a select group drawn from the total class as a sample. In this case the problem arises as to how far results based on such sample should be judged to be truly representative of the entire class of specifications (most of which have not been analysed). A problem of this nature may be called the problem of validity for this kind of work.

Such a validation problem appears to take the same form whether the product to be validated is a derivative specification or (derivative) check list. Accordingly we shall for the sake of brevity carry on the discussion by referring to the problem as that of validating (derivative) check lists. The same principles would apply if the product happened to be a derivative specification.

In order to consider the validity of a check list based on a sample group of specifications (called here a Sample Check List) we may hypothesize a check list based in the same manner on the entire class of specifications from which the sample was drawn. Such a hypothetical check list (which is not made) will be called the Ideal Check List. Then the problem of validity may be conceived as the question as to how far the content of the Sample Check List agrees with the unknown content of the Ideal Check List.

An overlapping of the two appears ordinarily to be certain but a failure of complete coincidence is very highly probable. The question is what degree of coincidence is to be expected.

This general validity problem naturally divides into two separate questions. The first question asks what proportion of the content of the Sample Check List may be expected to be present also in the Ideal Check List; this may be called the (sub-) problem of reliability. The second question asks what proportion of the content of the Ideal Check List may be expected to be present in the Sample Check List; this may be called the (sub-) problem of completeness. The answers to these two problems, if expressed in numerical percentages, could be called the Index of Reliability and Index of Completeness respectively.

We shall first consider these two problems in their simplest form and afterward in a more complex form in which they exhibited themselves in a recent study by the writer.[3] The simple case presents no great difficulty and it is possible that a different method of disposing of it might be preferred. The more complex case, however, appears to be rather difficult of solution and the writer has not been able to find in the literature any developed technique for handling it. The simple case is presented here primarily because it affords, by further extension, a successful approach to the difficult problem of the more complex case.

Simple Case

Terms and Symbols

The "class of specifications" will be understood to consist of all specifications which belong to the whole class of specifications regarded as a source, a class which we claim to represent in our final product. In this problem the "class" will not be regarded as indefinitely large but as consisting of a definite number of specifications, a number to be ascertained by actual count or by careful estimate.

"Sample specifications" are the limited group selected from the class for purposes of actual analysis, and which play the rôle of representing the whole class. The remaining specifications of the class are not analyzed.

"Sample Check List Material" is a name for the assemblage of all the different items found in one or more sample specifications.

"Ideal Check List Material" is a name for a hypothetical assemblage of all the different items found in one or more specifications in the class. Only those appearing in some sample specifications can be actually known, the rest are hypothetical.

[3] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

Write

- M (constant) = total number of specifications in class
- N (variable) = number of these specifications in which a particular item under consideration appears (this number is hypothetical and some of the particular items themselves are hypothetical)
- m (constant) = number of sample specifications
- n (variable) = number of sample specifications in which a particular (the same) item appears

Values of n may be expected to vary for different items, from m to 0 by intervals of 1, the zero value appertaining to any item wholly absent from the Sample Check List Material (hypothetically present in Ideal Check List Material).

Values of N might be expected to vary, for different items, from M to 1 by intervals of 1. But in this problem the convention will be adopted that the range is from M downward by intervals of $\frac{M}{m}$. Thus if the number M should be five times as large as the number m then the range for N would be treated as proceeding from M downward by intervals of 5: $M, M - 5, M - 10, \cdots 5$.

A "tabulation" will mean a statistical table showing how many different items appear in every possible number of specifications. A tabulation must be made by actual count for the items of the sample specifications, and will show the number of items having each possible value of n. A similar tabulation is hypothetical for the items in all the specifications of the class, that is for the number of items having each value of N permitted by the convention of the last paragraph.

"Tabulation cell" (or simply "cell") will mean, as needed, either the number of items or the group of items appearing in any designated number of specifications. For Sample Check List Material it will be the number or group of items to which a particular value of n appertains; for Ideal Check List similarly the number of items or group of items to which a particular value of N appertains (hypothetically).

"Sample Check List" will mean a list of items selected from the Sample Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be, for example, the numerical ratio $n \geqq \frac{m}{2}$.

"Ideal Check List" will mean a list of items selected from the Ideal Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be the numerical ratio $N \geqq \frac{M}{2}$.

Problem of Reliability

The problem of reliability may be restated and renamed the General Reliability Problem. This may be broken up into a group of problems which will be called Elementary Reliability Problems. Each of the latter may be in turn broken up into a group of problems which will be called Ultimate Reliability Problems. Each Ultimate Reliability Problem may be solved directly. Combination of these solutions will yield solutions of the Elementary Reliability Problems. Combinations of the latter solutions will finally yield the solution of the General Reliability Problem.

These problems will now be stated

General Reliability Problem: What proportion of the items present in Sample Check List may be expected to be present also in Ideal Check List?

Elementary Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in Ideal Check List?

Ultimate Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in some designated cell in Ideal Check List?

To solve an Ultimate Problem:

From the Fundamental Theorem in the Theory of Inductive Probability (Whittaker, E. T. and Robinson, G. The Calculus of Observations. London: Blackie & Son. 1924. p. 305) the solution may be expressed as

$$\frac{P_R \cdot p_s}{\Sigma\, Pp}.$$

Whittaker and Robinson's statement of the Fundamental Theorem in the Theory of Inductive Probability is as follows (form slightly changed without change in meaning):

"Suppose that a certain observed phenomenon may be accounted for by any one of a certain number of hypotheses, of which one, and not more than one, must be true: suppose moreover that the probability of the R-th hypothesis, as based on information in our possession before the phenomenon is observed, is P_R, while the probability of the observed phenomenon, on the assumption of the truth of the R-th hypothesis, is p_s. Then when the observation of the phenomenon is taken into consideration, the probability of the R-th hypothesis is

$$\frac{P_R \cdot p_s}{\Sigma\, Pp}$$

where the symbol Σ denotes the summation over all the hypotheses."[4]

It is clear that an Ultimate Reliability Problem is a case falling under this Fundamental Theorem. The observed phenomenon is any item occurring in any specified cell of Sample Check List, say cell $n = s$. It may be accounted for by a certain number of hypotheses as to its source in the Ideal Check List

[4] For the fundamental position of this theorem in a theory of science and for its proof one may also consult Jeffreys, H. Scientific Inference. Cambridge: Cambridge University Press. 1931. Chapter II (section 2.34).

Material; the different cells in the Ideal Check List Material are these different hypotheses of origin, hypothetical because we do not *know* from which one it has come but only that it must have come from some one of them; the cell from which it actually comes is the true hypothesis, though we do not know which one that is. That the origin of the item is in cell $N = R$ is the R-th hypothesis, and its probability is written P_R. The probability of the occurrence of the phenomenon on the assumption of the truth of the R-th hypothesis is the probability that an item in cell $N = R$ will appear in Sample Check List in cell $n = s$ and its probability is written p_s. As we clearly have in our Ultimate Reliability Problem a case falling under the Fundamental Theorem quoted we may accept as the required solution of the Ultimate Reliability Problem the formula already given in the initial statement:

$$\frac{P_R \cdot p_s}{\Sigma\, Pp}.$$

This expresses the probability that any item found in Sample-Check-List cell $n = s$ comes from (and appears in) Ideal-Check-List-Material cell $N = R$, or it gives the proportion of items found in Sample-Check-List cell $n = s$ that may be expected to come from (or appear in) Ideal-Check-List-Material cell $N = R$.

Meaning of any value of P (say P_R) = the probability that any item, drawn at random from those cells of Ideal Check List Material which are possible sources of items in Sample-Check-List cell $n = s$, will happen to be drawn from cell $N = R$.

Meaning of any value of p (say p_s) = the probability that any item in Ideal-Check-List cell $N = R$ will also be present in Sample-Check-List cell $n = s$. (Important: this supposition is *not* equivalent to its converse.)

Evaluation of P_R:

$$P_R = \frac{\text{number of items in cell } N = R}{\text{number of items in all cells which are possible sources of items in cell } n = s}.$$

For this ratio it is necessary to assume that the shape of the numerical curve formed by the group of Ideal-Check-List-Material cells is the same as that of the numerical curve formed by the group of Sample-Check-List-Material cells. On this assumption we may replace the numerator by the number of items in the Sample-Check-List-Material cell having an abscissa corresponding to that of the Ideal-Check-List-Material cell $N = R$, and replace the denominator by the sum of the numbers of items in all the cells with abscissae corresponding to those of Ideal-Check-List-Material cells which are possible sources of items in cell $n = s$.

Evaluation of p_s:

By the aid of "the definition of probability which is used in practically all treatises on the subject" (Coolidge, J. L. An Introduction to Mathematical

Probability. Oxford: Oxford University Press. 1925. p. 4) and the principle underlying the Theory of Combinations (Whitworth, W. A. Choice and Chance. New York: G. E. Stechert & Co. 1927. Proposition II) we are able to arrive at the evaluation:

$$p = \frac{C_{m-n}^{M-N} C_n^N}{C_m^M}$$

in which, for any p (say p_s), we employ for N the value $N = R$, and for n the value $n = s$. As the denominator later cancels out it may be disregarded throughout, simplifying the formula to

$$p = C_{m-n}^{M-N} C_n^N.$$

(A symbol such as C_n^N is read "the number of combinations of N things taken n at a time"; also written in several other forms.)

The definition referred to may be worded as follows (Coolidge's own preferred definition is not quite the same):

"An event can happen in a certain number of ways, which are all equally likely. A certain proportion of these are classed as *favorable*. The ratio of the number of favorable ways to the total number is called the probability that the event will turn out favorably."

The principle underlying the Theory of Combinations may be quoted from Whitworth as follows (also found in ordinary works on algebra):

"If one operation can be performed in m ways, and then a second can be performed in n ways, and then a third in r ways, (and so on), the number of ways of performing all the operations will be $m \times n \times r \times$ etc."

If it is not at once clear that the formula for evaluation of p follows from the definition and principle just quoted, the following considerations should make it evident.

We are working in terms of a particular item belonging to a particular Ideal-Check-List-Material cell, say cell $N = R$. "Favorable" occurrence requires that this item fall in a particular Sample-Check-List cell, say $n = s$, while falling in any other Sample-Check-List-Material cell (including cell $n = 0$ for absence) is "unfavorable." Again the real meaning of the "favorable" occurrence is that the item will be found in just $n = s$ out of the m specifications of the sample, and absent in the remaining $m - n$ specifications of the sample. Moreover presence in Ideal-Check-List-Material cell $N = R$ means that the item occurs in just $N = R$ of the M specifications that constitute the whole class and is absent in $M - N$ of these specifications. The total number of all the ways (favorable and unfavorable) in which our event can happen means the same as the total number of all the ways in which a group of m specifications can be selected from a larger group of M, and this is, of course, written C_m^M and given us in our denominator. The number of favorable ways in which our event can happen means the same as the number of ways in which N specifications containing the item can form groups of n specifications while at the same time $M - N$ specifications not containing the item can form groups of $m - n$ specifications; the first distribution can be done in C_n^N ways and the second in C_{m-n}^{M-N} ways, so by Whitworth's principle the number of ways which these things can happen simultaneously is $C_{m-n}^{M-N}\,C_n^N$. Assembling numerator and denominator we have the formula initially stated for evaluation of p, viz.:

$$p = \frac{C_{m-n}^{M-N}\,C_n^N}{C_m^M}.$$

This is the general formula; in applying to the particular example $N = R$, $n = s$ the replacements for N and n, of course, give

$$p_s = \frac{C_{m-s}^{M-R}\,C_s^R}{C_m^M}.$$

Having a means of evaluating P and p we may solve all needed Ultimate Problems. The resulting solutions of the needed Ultimate Reliability Problems (not necessarily completed) enables us to arrive at the solution of any needed Elementary Reliability Problem in the form of a percentage which may be called an Index of Reliability for the Sample-Check-List cell in question. In computing this percentage we distinguish source-cells that belong to the Ideal Check List from other source-cells that belong to the Ideal Check List Material but not to the Ideal Check List.

By properly averaging cell-Indices of Reliability (which are really Indices of Reliability for the individual items in the cells) we may obtain a solution of the General Problem of Reliability in the form of an Average Index of Reliability for the Sample Check List as a whole.

In addition to the Average Index of Reliability for *the* Sample Check List we may easily secure also Average Indices of Reliability for any series of briefer Sample Check Lists selected from *the* Sample Check List, by properly averaging the Indices of cells contained in any Sample Check List in question, keeping the original criterion for Ideal Check List.

In practice it may not be necessary to compute all cell-Indices, as a portion of these may be entered in tables by any methods of interpolation regarded as acceptable.

Problem of Completeness

Again we have General, Elementary, and Ultimate Problems. These may be stated as follows:

General Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in Sample Check List?

Elementary Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

Ultimate Completeness Problem: What proportion of the items in a particular cell in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

To solve an Ultimate Problem:

From principles already used the proportion to be expected is the same as the value of p alone in an Ultimate Reliability Problem, viz.:

$$\frac{C_{m-n}^{M-N} C_n^N}{C_m^M}.$$

By the use of this formula we may solve the Ultimate Problems for all values of N represented in Ideal Check List and all values of n represented in Sample Check List; some of these solutions will have a value of zero.

For each value of n, if we properly average the solutions of the Ultimate Problems, we obtain a solution of the Elementary Problem for one Sample-Check-List cell in the form of a percentage which may be called the Index of Completeness for the particular Sample-Check-List cell. In securing this average it is necessary to multiply each Ultimate Problem solution by a relative number corresponding to the assumed ratio of number of items in the particular Ideal-Check-List cell to the number of items in all the Ideal-Check-List cells. The source of the assumed relative numbers is the same as that used in evaluating P in the Reliability Problem.

When we have an Index of Completeness for each Sample-Check-List cell we may obtain a Total Index of Completeness for the Sample Check List as a whole by summing the cell-Indices of Completeness of all the cells of the Sample Check List. By an equivalent but preferable method we may divide the last-named result by the sum of the cell-Indices of Completeness of all the cells of the Sample Check List Material (including cell $n = 0$); by this method the C_m^M of the original formula cancels out and so may be disregarded throughout.

A Total Index of Completeness is similarly obtainable for *a* Sample Check List (any Sample Check List selected from *the* Sample Check List) by summing the cell-Indices of Completeness of the appropriate cells. Thus, if desired, a tabulation may be made showing Indices of Completeness for a series of Sample Check Lists differing in extent.

A combined tabulation may show for each of a series of Sample Check Lists its Index of Reliability and its Index of Completeness.

More Complex Case

So far we have considered a validation problem of simple type. In the writer's Check List Materials for Public School Building Specifications[5] a more complex problem was presented, due to the introduction of the concept of the Applicable Case. A Check List for School Building Specifications was developed with a view to its use by school officials or others as an aid in judging proposed school building specifications with reference to their completeness or incompleteness of determination. The position was taken that a new specification ought not to be charged with the omission of a given item unless the building (as repre-

[5] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

sented by the specification) had an Applicable Case for that item. To give a single example, the Check List contains various items relating to the specifying of marble work. It did not seem appropriate to score a specification down for the omission of numerous determinations in marble work, if in fact there was no marble in the building to be determined. This situation is expressed by saying that there are no Applicable Cases for those items.

It seems likely that there are other research problems in which the question ought to be raised whether adequate treatment does not require the introduction of the concept of the Applicable Case. If so a more difficult validation problem is presented than would otherwise be the case.

In the more complex case indicated solution is obtained by making the necessary extensions in the procedures followed for the simple case.

Modifications in Terms and Symbols

M (constant) = total number of specifications in class
D (variable) = number of these specifications containing an Applicable Case for a particular item
N (variable) = number of the latter specifications which also contain the particular item
m (constant) = number of specifications in sample
d (variable) = number of these specifications containing an Applicable Case for the particular item
n (variable) = number of the latter specifications which also contain the particular item

Values of d range from m to 0 by intervals of 1, and those of n range from d to 0 by intervals of 1.

The convention is adopted that values of D range from M downward, and those of N from D downward, by intervals of $\frac{M}{m}$.

(Tabulation) cell will mean the number of items (or the group of items) having a common value of d and a common value of n.

The criterion for membership in the Sample Check List may, for illustrative purposes, be taken as $n \geqq \frac{d}{2}$.

The criterion for membership in the Ideal Check List may, for illustrative purposes, be taken as $N \geqq \frac{D}{2}$.

Problem of Reliability

Following the same principle and line of reasoning as for the simple case we arrive at the same general formula for the solution of an Ultimate Reliability Problem, viz.:

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

Meanings of values of P and p are the same as before except that cells must be described respectively in terms of n and d values instead of n values alone, or N and D values instead of N values alone.

P_R is evaluated in the same manner as before, using the new meaning of "cell."

For p_s the evaluation now becomes

$$p = \frac{C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N}{C_m^M}$$

which through cancellation may be simplified to the working formula

$$p = C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N.$$

The reasoning leading to the denominator C_m^M is unchanged and so this denominator itself remains unchanged. The numerator for the evaluation of p is altered to the extent shown by the consideration that, in producing "favorable" ways, we now have to do with the number of simultaneous possibilities of drawing n specifications from a group of N specifications containing a particular item, drawing $d - n$ specifications from a group of $D - N$ specifications which contain an Applicable Case for this particular item but do not contain this item itself, and of drawing $m - d$ specifications from a group of $M - D$ specifications which contain no Applicable Case for the item.

Problem of Completeness

Following the same principles and line of reasoning as for the simple case we arrive at the following formula for the solution of an Ultimate Completeness Problem:

$$\frac{C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N}{C_m^M}.$$

By suitable treatment bringing about cancellations the working formula may be reduced to

$$C_{m-d}^{M-D} C_{d-n}^{D-N} C_n^N$$

Techniques and Aids in Computation

The present paper is limited to an attempt to explain with adequate fullness the proposed theory of validation for derivative specifications and check lists, and space is lacking in which to exhibit techniques of actual computation. One specimen problem worked out in fairly complete detail, together with remarks on available aids in computation will be found in Appendix A3 in typewritten copies of the writer's "Check List Materials for Public School Building Specifications" on file in the Library of Teachers College, Columbia University; the Appendices are not included in the printed edition.