THE

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

Research and Studies

Volume 3
D–E

Editors-in-Chief

TORSTEN HUSEN

T. NEVILLE POSTLETHWAITE

THE INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

Research and Studies

Volume 3
D–E

Editors-in-Chief

TORSTEN HUSEN
University of Stockholm, Sweden

T. NEVILLE POSTLETHWAITE
University of Hamburg, FRG

PERGAMON PRESS
OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon Press Canada Ltd., Suite 104, 150 Consumers Rd., Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, Hammerweg 6, D-6242 Kronberg-Taunus, Federal Republic of Germany

First edition 1985

Library of Congress Cataloging in Publication Data
Main entry under title:
The International encyclopedia of education

Includes bibliographies.
Index: v. 10.
1. Education—Dictionaries. 2. Education—Research—Dictionaries. I. Husén, Torsten, 1916– II. Postlethwaite, T. Neville.
LB15.I569 1985 370'.3'21 84-20750

British Library Cataloguing in Publication Data
The International encyclopedia of education

1. Education—Dictionaries

I. Husén, Torsten II. Postlethwaite, T. Neville
370'.3'21 LB15

ISBN 0-08-028119-2

Computer data file designed and computer typeset by Page Bros (Norwich) Ltd.
Printed in Great Britain by A. Wheaton & Co. Ltd., Exeter

CONTENTS

Honorary Editorial Advisory Board	vii
Editorial Board	ix
Alphabetical Entries	Volumes 1–9
Classified List of Entries	Volume 10
List of Contributors	Volume 10
Author Index	Volume 10
Subject Index	Volume 10
List of Major Education Journals	Volume 10

D

Daily Living Skills

Daily living skills—sometimes called "survival skills" or "life skills"—include those competencies which are commonly used in day-to-day life. They are the capabilities which are necessary for functioning both at home and in society.

There are two worldwide educational trends which emphasize daily living skills. In countries where large portions of the population have had little or no formal education, educational programs are being expanded to provide more students with more instruction in the skills necessary for daily living. In countries where nearly everyone receives several years of schooling, educational programs are being changed by eliminating extraneous or very theoretical items from the compulsory curricula, to make room for a core of more practical or "relevant" competencies in a large number of subject areas.

There has been a tendency among curriculum planners in many countries (particularly since the emergence of competency-based education) to differentiate among three categories of core competencies: basic skills (providing a foundation for further education) (see *Basic Skills*), life skills, and job-preparation skills. There is, of course, a tremendous overlap between these categories, and most experts agree that numeracy and literacy are at the cognitive core of each.

1. Literacy

Literacy generally includes not only reading and writing, but also facility with the related language skills of speaking and listening. Even experts within individual countries disagree significantly about how developed these skills must be in order to achieve "minimal" competence. Nevertheless, they do recognize that literacy, at some level, is a prerequisite for normal, daily living in virtually every society. Furthermore, reading, writing, speaking, and listening skills are fundamental learning tools and are therefore essential for further education and continued intellectual growth beyond "minimal" or "survival" levels.

It is worth mentioning here that literacy has not always been recognized as a necessary skill for daily living. At the end of the eighteenth century, "literacy" was frequently judged by the capacity of a person to form the letters of their name. Furthermore, even that minimal level of competence was mastered by less than half of the population in the developed nations of Europe (Resnick and Resnick 1977). It is quite apparent that, over the last 200 years, there has been a substantial increase in societal expectations concerning literacy and a substantial increase in actual skill levels (by whatever definition of literacy used).

There seems to be general agreement that complex technological and/or bureaucratic societies demand higher levels of literacy. However, even in the rural areas of Asia, Africa, and Latin America, there is evidence of a growing determination to increase the level of literacy. The five-year readership promotion campaign in Malaysia (launched in 1980) and the earlier Brazilian literacy movement (MOBRAL) are just two examples of a worldwide trend.

2. Computer Literacy

An emerging "convenience" in a few highly technological societies, computer literacy may well become a necessity of survival in the not-so-distant future. Already, in the Federal Republic of Germany, for example, the Federal States, the *Gesellschaft fuer Informatik*, and various committees of the Ministry of Science have demanded that the basics of computers and informatics be taught in the schools. Similar suggestions can be heard from Scotland's Munn Committee (1977) and in the new curriculum (Lgr 80) implemented in Sweden in 1982.

In several countries computer literacy is regarded as a "daily living skill" for tomorrow's adults. It is therefore appropriately being identified as an area for emphasis in today's schools.

3. Mathematics

Curricular change in mathematics has been markedly tumultuous during the 1970s and early 1980s. There has been an almost worldwide shift in emphasis from the highly theoretical to the more practical—away from "new math" (see *New Mathematics*), and "back to basics" (see *Back to Basics Movement*). In other words, there has been an increasing educational focus on daily living skills, job-preparation skills, and basic skills needed for further education.

In the Federal Republic of Germany, a single term—*mathematische Allgemeinbildung*—is used to include all three of the above skill categories, and it is being given increased emphasis in the curriculum. In Sweden, Kenya, and South Australia, new mathematics curricula (implemented in the early 1980s) quite explicitly lay special stress on daily living skills. Each of these programs includes extensive testing to monitor the mastery of individual skills.

Numerous studies are being made to identify precise lists of necessary mathematical competencies. One of the most convenient sources of input is potential employers. In England, for example, the Shell Centre for Mathematical Education has been doing extensive research on the mathematical needs of school leavers entering employment. The Education for the Industrial Society Project has been performing similar investigations in Scotland, as has the University of Klagenfurt in Austria.

In some school systems, a focus on survival or daily living skills has been seen as a rationale for trimming down the curriculum and, concurrently, trimming down the budget. However, while serious inquiries are identifying some mathematical skills which do not have to be taught, they are identifying many more topics which are not being emphasized and should be. A 1979 survey by the Science Education Center of the University of the Philippines, and a slightly earlier study in the Federal Republic of Germany, are just two examples which support the growing realization that "trimming back" curricula to include basic, necessary, or relevant skills will result, in most societies, in an increase in the amount of mathematics instruction.

4. *Multiple Subject Areas*

While numeracy and literacy are widely thought of as the academic core of daily living skills, they are not generally considered the only skills necessary for survival in day-to-day life. Unfortunately, it would be quite impossible to construct a universally acceptable list of such skills. The individual items would vary significantly from country to country (depending on the level of technological development, the political system, cultural traditions, and even the climate). Nevertheless, in specific geographical areas, efforts have been made to identify (and implement in the school curriculum) a broad range of competencies used in daily living.

In British Columbia (Canada), for example, the Ministry of Education has introduced a core curriculum identifying the competencies which are generally accepted there as essential—competencies which should be mastered by all students throughout the province. This core curriculum focuses on basic skills in the areas of language, measurement and computation, scientific approach, cultural and physical heritage, analysis research, study and problem solving, and healthful living (Ministry of Education 1977).

The Maryland (USA) State Board of Education has outlined five areas in which students should develop at least miminum competencies (necessary for graduation after 1982): basic skills (numeracy and literacy), survival skills (consumer, parenting, interpersonal, mechanical, and financial), work skills, leisure skills, and citizenship skills (see *Basic Skills*; *Leisure-time Education*; *Civic Education*).

The important characteristic of the new curricula in both British Columbia and Maryland is the emphasis placed on learning skills which are needed for living, for working, and for learning. Similar movements toward educational relevance in multiple subject areas can be found at various stages of consideration or implementation throughout the world. In Scotland, the Munn Report (1977) provides one instance of curriculum change at the consideration stage. In Australia, the Northern Territory Department of Education reached the implementation stage in seven subject areas in 1981.

The relevance of education to daily living experiences has become a fundamental concern throughout the world. There is a clear understanding that, while in school, individuals must be given more knowledge and must develop more skills than they would have needed to survive a generation ago. Educational planning, then, has focused on providing the opportunities for more learning to more students, and on assuring that programs include the minimal competencies for daily living.

Bibliography

Avakov R M 1980 *The Future of Education and the Education of the Future*. UNESCO, Paris [also published in French]

Behrstock J 1981 Reaching the rural reader. *J. Read.* 24: 712–18

British Columbia Ministry of Education 1977 *Guide to the Core Curriculum*. Ministry of Education, Richmond, British Columbia

Resnick D P, Resnick L B 1977 The nature of literacy: An historical exploration. *Harvard Educ. Rev.* 47: 370–85

Riehs R J 1981 *An International Review of Minimal Competency Programs in Mathematics*. SMEAC Information Reference Center, Ohio State University, Columbus, Ohio

Scottish Education Department 1977 *The Structure of the Curriculum in the Third and Fourth Year of the Scottish Secondary School*. (Munn Report). Her Majesty's Stationery Office, Edinburgh

R. J. Riehs

Dalton Plan

The system of dividing the subjects of the curriculum into two parts and providing highly individualized contracts of work to students for the academic subjects, and class groups for the vocational, social, and physical activities developed by Helen Parkhurst was known as the "Dalton Plan". It was very popular and widely imitated throughout the United States, the Soviet Union, England and Wales, other English-speaking countries, and Europe from the early 1920s. Academic subjects were organized sequentially and students worked individually. The vocational subjects were grouped and students worked in classes in a nongraded way. The freedom allowed to individual children and the organization of teaching were said to increase the efficiency of schools when compared to the traditional forms of school organization and instruction, and to introduce into schools "community principles and practices". The Dalton plan offered procedures for organizing learning when changes were being contemplated, and appealed to the educational progressives' concern for freedom, individual expression, and social cooperation as an alternative to the formalism of the class lesson.

Helen Parkhurst developed the Dalton plan, known also as the "Dalton Laboratory Plan", to meet the criticisms of contemporary education and to provide a favourable environment in which children could prepare for life, freedom, and responsibility as "industrious, sincere, open-minded, and independent" individuals (Parkhurst 1923 p. 5). Learning had to be combined with experience to test character, to form judgment, and to develop self-discipline in social experience. She drew on her experience teaching in a rural school with 40 pupils over eight grades and in high school, primary school, normal training schools, and a training college. Her reading in 1908 of E. J. Swift's *Mind in the Making: A Study in Mental Development* (Scribners', New York) introduced the idea of an "educational laboratory" where student activity replaced the didactic method. A plan of work for children between 8 and 12 years of age which was finalized in 1913 "aimed at the entire reorganization of school life" (Parkhurst 1923 p. 11). Further work eliminated the restriction of the timetable through organizing pupils into groups with a free choice of the studies in laboratories with specialist instructors. Additional experience included working with Dr Maria Montessori in Italy and introducing the Montessori Method to California in 1915 (see *Montessori Method*), undertaking a practical test of the laboratory plan through the help of Dr F. Burk, and more work on the laboratory plan in 1918 with the support of Dr M. V. O'Shea of Wisconsin University. In 1919 the laboratory plan was applied in the ungraded Berkshire Cripple School for boys, and, after attracting much interest, it was introduced on a larger scale in 1920 at the Dalton High School, Massachusetts. One early visitor to the school was Belle Rennie, an English pioneer of educational change, whose account of the Dalton laboratory plan in the *Times Educational Supplement* in May 1920 led to Helen Parkhurst visiting England in July 1921. Her ideas were received enthusiastically.

The Dalton plan called for the reorganization of the school so that it functioned as a community where the individual was free to develop in culture and experience and prepare for life. The school both provided freedom for the students to work without interruption and at their own pace, and required cooperation in the social experience of the school community. Character and knowledge were determined by the experience of living and working as a member of society rather than upon the subjects of the curriculum. The Dalton plan was not advanced as a panacea for academic ailments. Rather it provided "a way through which the teacher can get at the problem of child psychology and the pupil at the problem of learning". School situations were diagnosed "in terms of boys and girls. Subject difficulties concern students, not teachers. The curriculum is but our technique, a means to an end" (Parkhurst 1923 p. 23).

Teaching and learning were reconciled in the educational reorganization involved in implementing the Dalton plan. Once the curriculum was agreed to, students were assigned to a class. The work for 12 months was presented to all students at the beginning of the year. Students accepted the tasks assigned for each month as contracts to be signed and returned to the teacher when the tasks were completed. The curriculum was arranged for convenience into major and minor subjects. The major subjects were: mathematics, history, science, English, geography, foreign languages, and so on. The minor subjects were: music, art, handiwork, domestic science, manual training, gymnastics, and so on. Students progressed at their own rate, organizing their methods of working as they thought best. This arrangement secured the understanding of the work and gave students a sense of purpose and responsibility.

One laboratory for each subject of the curriculum with a specialist teacher in each was an essential feature of the Dalton plan. The laboratories were places where the students were free to work on their contract tasks without the distraction of shifts from one task to another determined by a time-table. Group work was encouraged by the requirement that all members of any class in any laboratory at one time should work together as a stimulus to discussion and as part of the exercise of social influences. The progress of students was recorded on graphs and reviewed regularly by the teacher and the students as a vital means of assessing progress and providing support.

Written assignments were central to the contract system. These set out the work to be covered in each subject in as much detail as determined by the specialist teachers bearing in mind the books, equipment, and other teaching materials available in the relevant laboratory. Typically the school day was divided into free time for work on contract assignments in the morning from 8.45 a.m. to 12 noon, followed by a pupil assembly and faculty conference until 12.30 p.m., and group conferences for reviewing progress until lunch at 1.00 p.m. The afternoon session was devoted to work in class groups on vocational or recreational activities.

The widespread popularity of the Dalton plan lay, in part, in the ease with which it could be varied and modified to suit particular circumstances such as limitations of space and the size of schools. Parkhurst encouraged this provided the spirit of the plan was preserved and schools for children under 9 years of age were excluded. Despite the emphasis on the suitability of the Dalton plan in catering for individual differences, the prescription of monthly work contracts in the case of slow or reluctant learners became a major limitation on the success of the plan as outlined by Parkhurst. It also lacked the detailed form and preoccupation with research and experimentation of the Winnetka scheme.

See also: Curriculum Integration; Winnetka Scheme; Individualized Instruction; Keller Plan: A Personalized System of Instruction; Teaching Methods, History of

Bibliography

Connell W F 1980 *A History of Education in the Twentieth Century World.* Curriculum Development Centre, Canberra

Dewey E 1922 *The Dalton Laboratory Plan.* Dent, London

Parkhurst H 1923 *Education and the Dalton Plan.* Bell, London

Selleck R J W 1972 *English Primary Education and the Progressives, 1914–1939.* Routledge and Kegan Paul, London

J. R. Lawry

Dance: Educational Programs

Within the American public school system, dance has been traditionally associated with physical education. This is reflected in the place of dance in the curriculum, the training and certification of dance teachers, and the content of learning experiences at the elementary and secondary levels. Currently, the field is marked by conflicting perceptions as to the meaning and function of dance as a separate discipline, from the conventional wisdom of society without, and from the educational community within. Available research while sparse and unstandardized, presents initial data on the status of dance in the schools and its attendant problems.

Dance in education has its origins in the expressive, improvisational forms developed by Isadora Duncan in the early part of the twentieth century. Her approach to movement and the use of the body, as a dramatic departure from ballet, was embraced by some educators as a unique contribution to the development of students. Its pertinence was viewed as evoking in students a consciousness of themselves as individual entities at a time of increasing mass education, allowing for their participation in an educational process based on their own personalities. This approach became a new current in dance education along with the prevailing folk dance and, on a more professional level, ballet. As most practitioners were involved in private schools, and dance (as folk dance) had become institutionalized into public schools, "modern" dance, as it was called, played a greater or lesser role in the ensuing years. Its entry into the public schools has been essentially through the efforts of individual teachers wishing to effect their own classes, or those few within the educational structure who recognized the arts as fundamental to education.

1. Definition

As an art form, dance may be defined as the expression of ideas, feelings, and sensory impressions in movement forms achieved through the unique use of the body. It is the language of movement which speaks through the vocabulary of space, time, and force, that is, a movement is shaped in relation to the space it occupies, the time it uses, and the energy which gives it power. These elements constitute the materials of dance and are essential in forming its kinesthetic–visual image (Dimondstein 1971).

Movement comes into being only through the combined use of these elements, but there is no objectively defined sequence. Yet, each achieves definition through an investigation of its formal properties: space, through direction, level, and range; time, expressed through the internal rhythms of the body (pulse, heartbeat, breath) and the external metrical rhythms (beat, accent, measure); force, through contrasts of sustained, swinging, percussive, vibratory movements.

2. Distinguishing Characteristics

Within the context of general education, the practice of making dance an adjunct to physical education has tended to equate it with recreation or physical skill, neglecting its characteristics as an art form. Although body control is the basis of all motor activity, control in dance differs from skills or techniques associated with sports or gymnastics. Dance is geared neither toward the refinement or skills in themselves nor toward competitive ends. Whether it is performed as an individual or group activity, its means are not rule bound as in a game, nor are its aims toward predetermined goals.

Dance involves kinesthetic perception, that is, an understanding and appreciation of movement developed through a "muscle sense." Evidence of such perception comes through a conscious awareness of the body: moving through space alone, in relation to others, and to the physical environment; responding to the dimensions of time, both metrical and created; resisting or acquiescing to gravity by restraining or expending energy. All of these function toward the controlled use of the body in expressing ideas and feelings. Thus, while work in dance techniques aids students in developing motor skills, technique alone is not sufficient in this process of discovering the qualities of expression that accompany each movement pattern.

The progression is one of transforming natural or basic movements (walking, running, leaping, swinging, turning) into improvisations (using basic movements to respond spontanecusly to specific ideas intitiated by the student or teacher), into dance studies (arranging movements into more organized sequences), into choreography (structuring more complex dances using conventional or invented forms). In addition to the performing aspects, a well-balanced program helps students at all levels to develop critical judgments of their own work and that of others, and to communicate their ideas in appropriate verbal terms. The process may not be sequential and is not determined by age. It is one of presenting creative dance problems which move from simple to increasingly sophisticated solutions, to which students respond according to their sensibility and capacity at any given time.

3. *Research*

There was little research until the 1960s, a period of general curriculum assessment. Educators, aided by federal money, reconsidered the role of the arts in affecting general education. The perceived dichotomy between dance as a performing art and as an educational discipline led to innovative projects such as "Arts in General Education" (dance related to other subject matter) (Madeja 1973), "Aesthetic Education" (dance as a separate area), and "Artists in Schools" (professional dance residencies) (National Endowment for the Arts 1976). Most existed on a regional basis or were designed for particular school districts; some were initiated by dance educators within school systems through extracurricular local agreements. Evaluations, while programmatic, revealed a range of student achievements from dance used as a unique mode of knowing and feeling, to incremental learnings in other subjects, to a heightened interest in schooling. Such implications, however, did not penetrate the mainstream of general education.

Limited research within the field reflects internal contradictions both theoretical and practical. A 1967 study conducted to determine the status of children's dance in elementary schools revealed fundamental misconceptions among practitioners as to the nature and function of dance (American Alliance for Health, Physical Education, and Recreation 1971). Although termed "dance," activities fell mostly into categories of rhythmic games, calesthenics, folk and square dance (with major emphasis on the latter). Curriculum leaders were found lacking a rationale as to the contribution of dance to the total development of young children, resulting in no clarity of purpose or coherent design. Problems were identified as inadequate college preparation, lack of inservice programs, and insufficient interest from sports-oriented personnel. Most pervasive is the conclusion that dance is relegated to the fringes of physical education programs and that even when taught in relation to other subjects, it retains a secondary character.

A survey on the status of dance in the secondary schools concerned the training and certification of teachers, and opportunities for students (American Alliance for Health, Physical Education, and Recreation 1971). The findings confirmed the previous study that college programs consisted primarily of physical education courses; some offered dance courses within physical education, a few included dance majors or minors, and almost none provided graduate degrees in dance education. In terms of opportunities for students, dance classes were offered in junior high (ages 12 to 14) with the number and frequency increasing in senior high (ages 15 to 18), with a maximum at age 15. Lacking specialization, most teachers had physical education backgrounds and were found inadequately trained in dance. Significantly, dance classes at this level are available only as "electives" for talented or interested students and do not qualify with other subjects for college entrance (see *Elective Subjects*).

Thus, available research suggests that dance is not regarded as a substantive area within general education, and is not given academic parity in terms of professional preparation, instruction, time, and facilities. Despite the successes of many experimental projects, most states still do not require dance as part of the curriculum and offerings are largely determined by the size of the school population and availability of competent staff. Contributing factors are school administrators unresponsive to granting dance a higher status, curriculum developers with insufficient training and experience to conceptualize programs, and lack of state certification to standardize requirements.

4. *Issues and Concerns*

Apart from empirical research, the controversies continue to generate literature around two central issues: the lack of consensus between physical education and dance education regarding the place of dance in the curriculum, and its function in the larger educational spectrum.

Among physical educators there are several currents, the dominant of which accepts no separation of dance from physical education, but projects a comprehensive program in which both would be complementary rather than competitive. The position taken by dance educators (and some generalists) is that dance is a form of educational development not available in other academic subjects. As such, it is a body of knowledge which requires study of its historical, aesthetic, and performing aspects, as well as its connections to other areas. Therefore, the important issues to be examined are the aims and content of dance as the theoretical framework for curriculum planning and research. These issues project a curriculum geared toward the development of aesthetic perception of dance as an art, that is, an ability to understand the formal, kinesthetic, and expressive elements found in various forms, and to interpret these meanings in relation to other knowledge. Such learnings would be acquired through direct participation, observation of performances, and reading and writing about dance.

To establish dance as a basic part of the school program would be to focus on its distinctive content and its relation to the broad aims of education. The concern is that dance should be separated from physical education and that its relation to other areas of subject matter be one of coexistence. Thus, the position within physical education is one of doing what is being done better; the other implies a fundamental philosophical and organizational change in conception, leadership, and instruction.

See also: Physical Education Programs

Bibliography

American Alliance for Health, Physical Education, and Recreation (AAHPER) 1971 Task force, children's dance, dance division. *JOHPER* (Journal of Health, Physical Education, Recreation) June 1971, Reston, Virginia

American Alliance for Health, Physical Education, and Recreation (AAHPER) 1971 The status of dance in the secondary schools. *JOHPER* October 1971, Reston, Virginia

American Alliance for Health, Physical Education, Recreation and Dance (AAHPERD) 1981 *Children's Dance*. Reston, Virginia

Dimondstein G 1970 Space-time-force: An aesthetic construct. In: Haberman M, Meisel T (eds.) 1970 *Dance: An Art in Academe*. Teachers College Press, New York

Dimondstein G 1971 *Children Dance in the Classroom*. Macmillan, New York

Gray V, Percival R 1962 *Music, Movement, and Mime for Children*. Oxford University Press, London

Madeja S S 1973 *All the Arts for Every Child*. Final Report on the Arts in General Education Project in the School District of University City, Missouri. The JDR III Fund, New York

National Endowment for the Arts, 1976 *Artists in Schools*. National Endowment for the Arts, Washington, DC

Russell J 1968 *Creative Dance in the Primary School*. Praeger, New York

G. Dimondstein

Data Analysis: Exploratory

Exploratory data analysis (EDA) is a collection of specialized tools and an approach to the analysis of numerical data which emphasizes the use of graphic displays and outlier resistant methods to detect and model patterns in data. Numerous researchers and statisticians have contributed to the development of EDA but the primary source of ideas is generally acknowledged to be John Tukey. Although many EDA tools have been known for some time, Tukey has created new procedures, improved older ones, and knitted them all together into a systematic method. Tukey's work, only partially described in his book, *Exploratory Data Analysis* (Tukey 1977), provides the data analyst with new capabilities for uncovering the information contained in numerical data and for constructing descriptive models.

Data exploration, as Tukey envisages it, is not simply an exercise in the application of novel tools. It is a phase of the empirical research activity, one which follows data collection (or acquisition) and precedes the application of confirmatory or "classical" inferential procedures (Tukey 1973). It is, thus, part of that twilight zone which experienced researchers find so exciting and challenging, novice researchers fear and misunderstand, and few researchers ever report. The excitement of this phase of research derives in large measure from the prospect of discovering unforeseen or unexpected patterns in the data and, consequently, gaining new insights and understanding of natural phenomena. The fear that novices feel is partly a response to this uncertainty, but it is also partly due to traditional teaching which holds that "playing around" with data is not "good" science, not replicable, and, perhaps, fraudulent. Many experienced researchers pay lip service to this view, while surreptitiously employing ad hoc exploratory procedures that they have learned are essential to research. Exploratory data analysis by making exploration routine and systematic, and by using theoretically justifiable procedures, opens the exploratory phase of research to public review, enhances its effectiveness, and allows for replicability.

Because Tukey's methods exploit the natural behavior of measurements, they allow researchers to rely on their intuitions. The simple logic of the methods helps clarify the process of modeling data and, consequently, makes it easier to detect errors in data or departures from underlying assumptions. Much of this is due to the graphical devices Tukey invented which are central to this approach because of their ability to portray a wide range of patterns that data can take. Well-designed graphics, such as those used in EDA, are useful for the guided searching that characterizes exploration and are also attractive mechanisms for communicating results to nontechnical audiences. As a consequence, EDA can serve in data analysis and for reporting the results of an analysis.

Many of the methods in EDA fall on the frontiers of applied statistics. Two important topics in statistics today are the robustness and resistance of methods, terms which refer to the ability of a procedure to give reasonable results in the face of empirical deviations from underlying theoretical assumptions. Clearly, robust and resistant methods are particularly advantageous in social science research because empirical social science data are so often obtained in an ad hoc fashion, frequently under nonreplicable circumstances, on opportunistically defined variables whose relation to substantive theoretical constructs are vague at best. Exploratory data analysis is especially important in educational research, where many of the variables studied and data collected are brought into analyses not because well-verified, substantive theory demands their inclusion, but rather because investigators "feel" they ought to be, because they are "convenient" to use, or because measurements have been recorded in some assumed "reasonable" manner. Nor are the data typically produced as a consequence of a scientifically designed experiment. It is precisely in such research that EDA can be used to its greatest advantage because it is here that an open mind is an absolute necessity: the analyst rarely has the support of theoretically based expectations, and the real task confronting the data analyst is to explore—to search for ideas that make sense of the data (Simon 1977).

In the following brief description of EDA, only a few of the more usable techniques and the philosophical essence behind EDA are presented. Mathematical details are avoided but references to more extensive treatments are provided. The general objective of the procedures presented can be easily summarized. The procedures are tools for achieving resistant estimates of parameters for traditional additive and linear models. In this respect, they speak to a common empirical problem, the presence of outliers in data and the sensitivity of traditional methods of parameter estimation to highly deviant observations. Resistant analogs to three cases are presented: (a) a set of observations on a single factor at one level, (b) a set of observations on a single factor with multiple levels, and (c) a set of observations on two factors. In each case, the traditional approach to parameter estimation is mentioned first and then the EDA approach is detailed.

1. Organizing and Summarizing Individual Batches of Data

One of the first tangible products of a quantitative research project is a set of numbers, "data" that might contain information about the phenomenon or process under investigation. In many cases, the sheer amount of data to be analyzed can be overwhelming, leading an investigator to rely on summaries rather than dealing with all the values obtained. In addition to the impact that quantity can have, computer routines often present data values and summary statistics in a printed format which obscures rather than elucidates data properties. Automatically produced by routines designed to handle a wide variety of situations, output listings typically contain much that is distracting (e.g., decimal points, leading and trailing zeros, scientific notation) and little that is fundamental. In addition, such routines are usually designed to present values in what might be called an accounting framework, one that facilitates the location of identified values but provides little insight into the overall behavior of the data.

Even a small collection of data, for example, three variables for 50 cases, is extremely hard to visualize or to get a feel for. What is needed is a technique that preserves the detail of values but eliminates distracting noise and contributes to a first level of understanding. The stem-and-leaf display and the box plot are two such techniques.

2. Visual Organization: Stem-and-leaf Display

The stem-and-leaf display is an immensely useful and easily appreciated exploratory tool which can provide insightful first impressions. It combines the features of a sorting operation with those of a histogram. The basic procedure can be used to organize and provide information on a single batch of values in a wide variety of circumstances. (A batch is a collection of observations on a variable. The term is not meant to convey any notion of sampling from a population.)

Figure 1 presents a stem-and-leaf display of the

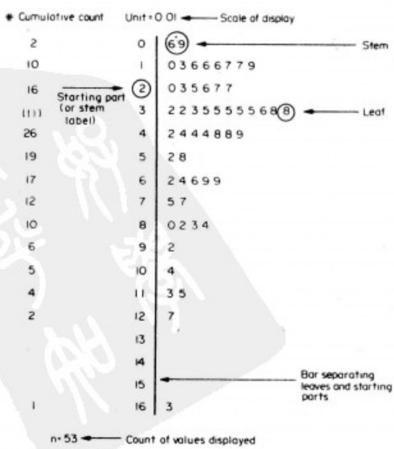

Figure 1
A stem-and-leaf display of direct silent

number of 5-minute segments out of 40 in which each of 53 children were observed to be reading silently and is referred to as direct silent in the figure (Leinhardt et al. 1981). The arrows, words, and circles are for explanation only. To construct a stem-and-leaf display, each number in a batch is partitioned into a starting part (to the left of the bar) and a leaf (to the right of the bar). When single digits are used in leaves each starting part will be an order of magnitude larger than each leaf. A set of leaves for a given starting part is a stem. The unit of display records the scaled value.

To reconstruct a data value, juxtapose a starting part with a leaf and multiply by the unit. For example, consider the two leaves that form the first stem of Fig. 1, 6 and 9. To reconstruct the two data values that these leaves represent, simply juxtapose each leaf with the common starting part, 0, and multiply by 0.01, that is, $06 \times 0.01 = 0.06$; $09 \times 0.01 = 0.09$. As another example, consider the bottom-most stem in Fig. 1. It has only one leaf, 3. Juxtaposing the 3 with its starting part, 16, and mulitplying by 0.01 yields 1.63. There are three starting parts 13, 14, and 15 that have no stems or leaves. This indicates that no observations have values between 1.27 and 1.63.

The display in Fig. 1 is actually the result of a two-step procedure (assuming the operation is carried out by hand). The first step normally yields a display in which the starting parts are ordered but the leaves on each stem are not. In the second step, each stem's leaves are ordered. This two-step procedure makes sorting a reasonably efficient operation.

Because all values in the display are represented by leaves occupying equal amounts of space, the length of a stem is proportional to the number of observations it contains. Thus, the display is like a histogram and provides information on shape and variation while also retaining information on individual values. This is true after the first step in construction. After the second step, the values are completely ordered and the display takes on the features of a sort. Because the display is like a histogram, anyone studying it can get the same kind of feeling for such elementary batch characteristics as overall pattern, bunching, hollows, outliers, and skewness that histograms provide. Those features that are akin to a sort allow the determination of maximum and minimum values quickly and, from them, the range of the values which can be used as a measure of overall variation.

Adding an inwardly cumulating count (depth) to the display greatly expands its utility. It facilitates finding other order statistics besides the maximum and minimum, such as the overall median and the medians of the upper and lower halves of the batch, which Tukey calls the "hinges." To form such a count, the number of leaves on a stem are cumulated from both ends in towards the middle. The median is located (not its value) at a depth halfway into the batch from either end. The count of the number of leaves on the stem containing the median is given and put between parentheses because it is not cumulative.

To illustrate the use of this column of inwardly cumulating counts, the count column will be used to find the values of the median of the data in Fig. 1. The median will be located at depth $(n + 1)/2$. Since there are 53 values, the median is at depth 27, that is, it is the 27th value in from either the high or low end of the sorted values. Counting into the batch from the low-value end (which happens to be at the top of this display), it can be seen that the 27th value is represented by a leaf on the fourth stem. The value of the median could just as easily have been determined by counting into the batch from the high-value end (at the bottom of the display).

While the stem-and-leaf display is useful for describing data, it can also be an effective exploratory tool. For example, looking at Fig. 1 an asymmetry can be seen skewing the values toward the high end. The clustering between 0.1 and 0.4 is obvious, as are the two groups at 0.6 and 0.8 and the modal group at 0.3. The minimum value, 0.06, and the maximum value, 1.63, are easily determined. There is a gap apparent between 1.27 and 1.63. A researcher might be concerned, even at this point, with the question of why the maximum value seems to straggle out so much.

3. *Numeric Summarization: Number Summaries and Letter-value Displays*

While the stem-and-leaf display is a convenient and easily understood tool, it has its drawbacks. This is most evident when different batches of values are being compared. Although a simple comparison of the shapes of two batches can be achieved by placing the leaves of one batch on one side and the leaves of the second batch on the other side of a common set of starting parts, simultaneously comparing three or more batches using stem-and-leaf displays is obviously going to be difficult, possibly even confusing. While the visual quality of the stem-and-leaf display is a true asset in any first look at the behavior of a batch, it may be burdensome to continue to work with all the data values at once rather than a set of summary statistics.

The question is which summary statistics to use. The problem with choosing the mean and related statistics, such as the standard deviation, is their lack of resistance to the impact that one or a few deviant data values can have. Because the mean is a ratio of a sum to the number of values making up the sum, it can be equal to anything by simply changing a single value in the sum. This is not a problem, of course, if the data are reasonably well-behaved. Empirical data, however, often contain deviant values. Indeed, "weird" or "funny" values are rather commonplace occurrences (recall the value of 1.63

in Fig. 1) and, regardless of their source, they can cause traditional summary statistics to misinform.

Other statistics exist which are less sensitive to deviant values than is the mean, and, while they may not yet be fully supported by the inference procedures available for the mean, they may still be preferable at the exploratory stage of an analysis, where inference is not yet a focal issue. Some of the more useful and commonly known resistant measures of location and variation can be derived from the median and other order statistics. Most order statistics are little affected by the presence of a few outliers in a batch. One common resistant order-statistic-based measure of variation is the interquartile range.

Tukey exploits the resistance of order statistics, especially the median, in EDA. His first step in the numerical summarization of a batch for exploratory purposes involves computing five order statistics: the median, the extremes (or maximum and minimum), and the medians of the upper and lower quartiles (i.e., the hinges). When these five numbers are grouped together, they are called a "five-number summary" and can be arrayed conveniently as LE(LH, M, UH)UE, for lower extreme, lower hinge, median, upper hinge, and upper extreme, respectively. Tukey has introduced a truncation rule to avoid the inconvenience of small fractional ranks when finding medians of segments of a batch. The rule is

$$\text{Depth of next median} = (1 + \lfloor \text{depth of prior median} \rfloor/2 \quad (1)$$

The symbols $\lfloor$ and $\rfloor$ refer to the mathematical "floor" function which returns the largest integer not exceeding the number. That is, the fractional component of a value, in this case a fractional depth or rank, is discarded. This means that the only fractional depths used will be those that lie halfway between two consecutive values and, thus, will be easy to compute and understand. As a consequence of this truncation rule, exploratory summary statistics may not be exactly equal to analogous-order statistics whose computation is derived from more mathematically precise definitions.

The notion of a median is easily extended to provide a way of segmenting a batch resistantly. The hinges are themselves medians of segments, the upper and lower halves of the batch. Medians of the upper and lower quarters halve the quarter so that each segment bounds one-eighth of the values; medians of these segments bound 16ths, then 32nds, then 64ths, and so forth. In EDA, this process works outward from the center to the edges of a batch providing more and more detailed information on the behavior of the tails of an empirical distribution.

Although letter-value displays and five-number summaries (and extended number summaries in which medians of further foldings are recorded) provide useful, resistant information on location; their primary analytic use is in facilitating the computation of other features of batches. Differences of values provide information on spread or variation in a batch. For example, the range of silent reading data in Fig. 1 is computed by subtracting the lower extreme from the upper extreme: $1.63 - 0.06 = 1.57$. The range, however, is not a very resistant measure of spread. Obviously, it is very sensitive to deviant values when these appear as extremes, a common occurrence. A more reasonable measure of spread is the range between the hinges. Analogous to the interquartile range, it is called a "hingespread." The hingespread (which is symbolized as dH) of the silent reading data is $0.69 - 0.26 = 0.43$.

The hingespread is a statistic of central importance in elementary EDA. It is a useful tool in the search for values that deserve attention because they deviate from most values in a batch. This search can be started by computing another measure of spread, the "step," which is 1.5 times the hingespread. Using this quantity, one literally steps away from each hinge toward the extreme to establish another boundary around the central component of the data. These bounding values are called the "inner fences." Another step beyond these establishes the "outer fences." Note that the fences are not rank order statistics but are computed distances measured in the same scale as the values of the batch. Values that fall between the inner and outer fences are called "outside" values, and beyond these lie the "far outside" values. The two data values (or more if multiple observations occur at the same point) falling just inside the inner fences are called "adjacent values"; they are literally next to or adjacent to the inner fences.

It is useful to re-examine the stem-and-leaf display in Fig. 1 in light of the new information obtained on spread. In examining this display, it was noted that the data were evidently skewed out toward the high end. The numerical information on spread confirms this visual impression and suggests that one value at the high end deserves further attention. This value may be erroneous, or it may have been generated by a process different from that which generated the bulk of the values. Having identified a potential outlier, the problem of deciding what to do about it arises. If data are used which are made to appear highly asymmetric because of a few extreme observations, then it must be realized that many of the usual forms of inference, such as analysis of variance and least squares regression, will be strongly influenced by these few values. These procedures are not very resistant and, while removing values from empirical data should be done with utmost caution, the fact must be faced that unless omission of outliers is explored, fitted parameter values may describe the behavior of only a very small portion of the data. Replicability of findings in such situations is unlikely and generalizability is questionable.

The theoretical rationale underlying this approach to identifying outliers is not explicitly developed in Tukey's book on EDA. Some implicit support is available, however, by examining the properties of a normally distributed population in terms of EDA order-statistic-based measures. In a Gaussian or normal population, $\frac{3}{4}\, dH$ is approximately one standard deviation. Thus, 1.5 dH, a step, is approximately 2σ. Consequently, the inner fences, which are more than 2σ from the median, bound over 99 percent of the values of such populations. Observations drawn from a normal population that lie beyond the population's outer fences, which are an additional 2σ farther out, should indeed be rare.

4. *Schematic Plots as Graphic Summaries*

The quantities contained in number summaries and letter-value displays provide useful information on overall batch behavior. Most analysts, however, and certainly most nonspecialists, find that they can more easily appreciate the nuances of quantitative information when this information is displayed graphically. A schematic plot is an extremely useful graphic representation of the quantities contained in a number summary and, in fact, might well be considered a fundamental EDA summary device. It completely eliminates numbers (leaving them to a reference scale) and selectively obscures the data, drawing attention to some values and not others. Those values that are completely obscured are the values lying between the hinges, on the one hand, and those lying between the adjacent values and the hinges on the other. Attention is drawn by single marks to all values lying beyond the adjacent values. An example using the silent reading data appears in Fig. 2, which also shows the two other techniques that have been previously described.

Schematic plot	Number summary		Stem-and-leaf display (Unit = 10^{-2})	
	LE	0.06	0	6 9
			1	0 3 6 6 6 7 7 9
	LH	0.26	2	0 3 5 6 7 7
	M	0.38	3	2 2 3 5 5 5 5 5 6 8 8
			4	2 4 4 4 8 8 9
			5	2 8
	UH	0.69	6	2 4 6 9 9
			7	5 7
			8	0 2 3 4
			9	2
			10	4
			11	3 5
			12	7
			13	
			14	
			15	
*	UE	1.63	16	3

Figure 2
Schematic plot, five-number summary, and stem-and-leaf display for direct silent reading data

Several points about schematic plots are worth noting. In the basic schematic plot, the width (or height, depending on orientation) of the box enclosing the central section of the data is arbitrary. However, this dimension can be used to represent information on other aspects of the data, such as batch size and significance of differences between medians (McGill et al. 1978). Whereas vertically oriented schematic plots are traditional and visually appealing, horizontal orientations are more effective for computerized printing operations because they permit a standard width to be used for any number of plots. Although the schematic plot is a visual device like the stem-and-leaf display, it is not as detailed and, indeed, is explicitly designed to reduce the amount of information being displayed. A related but even more elementary display, the box plot (so called because it consists of simply the box portion of a schematic plot and "whiskers" or lines extending to the extremes) obscures all except those in the five-number summary. Even though schematics speak to the issue of shape and spread, they can be somewhat misleading if gaps or multimodality occur between the adjacent values. Consequently, it is not advisable to use schematics as substitutes for stem-and-leaf displays, but rather as adjuncts.

5. *Transformations*

Frequently, naturally occurring data are modestly or extremely skewed, or exhibit some other property that make the data not normally distributed. Tukey emphasizes the need to consider the monotone transformation $y = kx^p$, where y is the transformed value, x is the original value, k is a constant set to -1 when p is less than zero and 1 otherwise. (The constant, k, retains order in the magnitude of the values when the transformation is a reciprocal.) The procedures for determining p are worked out and presented elsewhere and will not be described here (Leinhardt and Wasserman 1978). A summary of transformations is given in Table 1.

Thinking in terms of rescaled data values rather than raw data values is by no means straightforward and, given the central role that power transformations play in EDA, it is important that their rationale and validity is fully appreciated. There are several ways of thinking about transformations. One involves realizing that the well-grounded confirmatory tools of standard inferential statistics make specific assumptions concerning model structure and error properties. In many common procedures these include assumptions about normality of error distributions, additivity in parameters and variables, constancy of error variances, lack of functional dependence between error variance and variable location, and lack of interactions. When these assumptions are invalid, the procedures lose some of their appealing qualities. Their use in such problematic situations can be misleading. Unless procedures that deal directly with the known features of the data are used, one must resort to mathematical modification that adjusts the values so that their properties fit the assumptions of the model and/or estimation procedure. Transformations of scale can often provide the modifying mechanism.

An alternative view is more metatheoretical. In it the theoretical development of the social sciences is seen to trail that of the natural sciences in the sense of not yet having a well-developed, empirically verified, axiomatic and deductive body of theory from which the appropriate scale and dimension for representing a theoretical concept in terms of an empirical variable can be determined. Dimensional analysis in physics is an example of the power inherent in disciplines where such well-developed theory exists. In its absence, analysts must often use variables measured in arbitrary scales or variables defined in an ad hoc manner. Rarely is there any good reason to believe that such measures come in a form best suited to modeling relationships. In the absence of an a priori theory that could specify a model, EDA provides tools to determine whether rescaling a variable will lead to a better analysis.

6. Modeling Data

In EDA, models for data consist of two parts: a part that uses a mathematical statement to summarize a pattern in the data and a part that summarizes what is left over. Each observed value can be decomposed into a part that is typical of the pattern, the "fit," and a part that is not typical of the pattern, the "residual." Tukey constructs a verbal equation to represent a general class of models where the decomposition is additive:

$$\text{Data} = \text{Fit} + \text{Residual} \tag{2}$$

Other models are not ruled out, but Tukey emphasizes the use of simple models because they are easily understood, are easily estimated, help reveal more complex features, and often provide a good first approximation to these complexities. Additionally, many other forms can be rendered in terms of Eqn. (2) through an appropriately chosen transformation. Consequently, Eqn. (2) plays a fundamental role in EDA.

The computational procedure is straightforward. The median is subtracted from each observed value. This yields a batch of residuals, that is, a batch of adjusted values indicating the amount by which each raw value deviates from the fitted central value. In terms of a horizontally formatted schematic plot, the computation of residuals is analogous to centering the raw data around their median and relocating the zero point on the horizontal axis so that this origin rests exactly on the median.

Table 1
Summary of transformations: roles, procedures, and failures

Data structure	Problem	Procedure	Failures
(a) Single batch	Asymmetry	Summary table (or equation)	Multimodality large gap
(b) One-way array	Spread heterogeneity	Diagnostic plot of log (dH_i) vs. log (x_i)	Inconsistency in dH_i
(c) Two-way array	Interaction	Diagnostic plot of comparison values	Idiosyncratic interactions
(d) Paired observations	Curvature	Slope ratios or equation	Nonmonotonocity
	Spread heterogeneity	(b above)	(b above)

Note: Sometimes the "correct" transformation is not well-approximated by kx^p for any "reasonable" choice of p

7. A Model for Multiple Batches

Most research projects are not performed for the purpose of fitting single parameter models or obtaining a single summary statistic such as the mean or median. At the very least, the simplest objective involves comparing several batches in an attempt to determine whether one batch differs from another and by how much. A second-order question involves deciding whether an observed difference is important. These questions are traditionally approached through the analysis of variance (ANOVA) using ordinary least squares (OLS) estimation procedures (see *Analysis of Variance and Covariance*). While OLS has estimable properties when special conditions hold (i.e., the parameter estimates are unbiased, consistent, and have minimal variance), some of these properties are lost when the conditions fail to hold. Such losses can result from the presence of a single outlier. An EDA-based approach which exploits graphical displays to detect data inadequacies is presented here which employs resistant measures in determining effects and provides a useful guide in obtaining a transformation that facilitates the use of classical procedures.

In classical ANOVA, the errors, ε_{ij}, are assumed to be normally distributed random variables with zero mean and constant variance. Thus, the sum of squares for batch effects and the errors are multiples of χ^2 random variables. As a consequence, F ratios can be formed to test zero effect null hypotheses and symmetric confidence intervals can be constructed.

An analogous EDA procedure is presented here; one that is more resistant to outliers than ANOVA but lacking distributional assumptions and, consequently, lacking inferential tests. The purpose is to provide a resistant analysis that can be used in exploration. Furthermore, the EDA procedure provides a useful mechanism for studying the problem of inconsistency of variation in the errors, that is, heteroscedasticity.

The EDA modeling procedure is similar to that pursued in a classical analysis except that common effect and batch effect are estimated by medians and, consequently, involve different arithmetic computations. The model represented by the verbal Eqn. (2) yields a "fit" that is applicable to all batch values. For multiple batches, this model can be further elaborated so as to distinguish a general or common effect across all batches and a set of individual batch effects that are confined within their respective batches. Thus, the general model becomes:

$$\text{Data value}_{ij} = \overbrace{\text{Common effect} + \text{Batch}_j\text{ effect}}^{\text{Fit}} + \text{Residual}_{ij} \quad (3)$$

Conceptually, the model represents each observed data value as a conditional response determined in part by imprecision, noise, or error. No specific assumptions are made about this last ingredient except that, taken as a batch, the residuals are devoid of an easily described pattern. The information they contain relates solely to the overall quality of the model in terms of its ability to replicate the observed values.

The computational procedure is straightforward. First, consider the hypothetical multiple-batch data set represented by the box plots in Fig. 3. The median of the pooled batches is identified as the "common effect." Next, subtraction is used to "extract" this common effect from all data values. The result is simply a new centering of the adjusted batch values around a new grand median of zero. Second, the individual batch effects are obtained by subtracting the grand median from the individual batch medians.

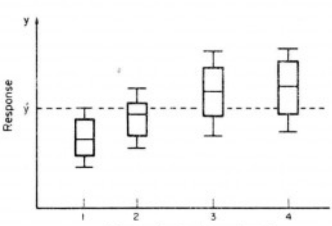

Figure 3
Box plots of multiple batches of hypothesis data

Finally, residuals are obtained by subtracting the batch effects from each adjusted value in the appropriate batch. The residuals are then examined as a whole and as batches. For the hypothetical example, the model is:

$$\text{Data value}_{ij} = \text{Common effect} + \begin{Bmatrix} \text{Batch effect}_1 \\ \text{Batch effect}_2 \\ \text{Batch effect}_3 \\ \text{Batch effect}_4 \end{Bmatrix} + \text{Residual}_{ij} \quad (4)$$

The fitted value for the i, jth observation would simply be:

$$\text{Fit}_{ij} = \text{Common effect} + \begin{Bmatrix} \text{Batch effect}_1 \\ \text{Batch effect}_2 \\ \text{Batch effect}_3 \\ \text{Batch effect}_4 \end{Bmatrix} \quad (5)$$

8. *A Model for Two-way Classifications*

A more complicated but quite common data structure arises when responses can be identified with the levels of two factors. The usual summary layout used to organize such data is the two-way table, an array of "responses" organized on the basis of row (r)

and column (c) factors. Such two-dimensional arrays consist of $r \times c$ cells or entries. Each row or column of a factor is referred to as a factor level or factor version. Factors are usually ordinally or nominally scaled but may be interval scaled. Responses are usually ratio scaled. The data are conceived of as triples of values: two classifying variables and a response variable.

The usual approach to such data involves an elaboration of the one-way model. A model, additive in factor-level effects, is posited. The array of responses is decomposed into an overall level or common effect, row effects, column effects, and interaction effects. A two-way ANOVA using least squares is the traditional method employed to estimate the model's parameters and to test for significance. In ANOVA, the grand mean is used to estimate the common term, and row and column means of the adjusted data estimate the row and column effects.

Once again, the EDA approach is analogous. The differences lie in the lack of distributional assumptions for the errors and the use of medians to estimate model parameters. Because no distributional assumptions are made, the hypothesis tests that are possible with least squares cannot be done. However, the use of medians ensures a result that is more resistant to the impact of deviant values. Furthermore, the EDA procedure provides a useful way to detect interactions even when there is only one observation per cell. When certain kinds of interactions are present, the EDA procedure can lead to a choice of a power transformation of the data that eliminates the interactions, that is, yields a scale in which the additive model provides a reasonable summarization of the data.

The model is in many respects an extension of the one-factor, multiple-level model proposed earlier for multiple batches. Indeed, a two-way table of responses can be thought of as two interwoven sets of multiple batches. Considering the column factor as the only dimension, there is a set of c multiple batches, each containing a maximum of r values. Considering the row factor as the only dimension yields a set of r multiple batches, each containing a maximum of c values.

Because the assumption is that each cell contains multiple observations, these data can be visualized as a two-way categorization of box plots as in Fig. 4. In the display, the vertical or y-axis is the scale on which the numeric response variable is measured. Factors R and C are categorical, so the distances between the levels, as well as their ordering, are arbitrary. The box plots of the multiply observed responses appear elevated above the origin plane by differing average positive amounts. The mathematical model represented involves a decomposition of the average elevation of each box plot into four parts: (a) an overall level; (b) a contribution from the column level (which occurs regardless of row level); (c) a contribution from the row level (which occurs regardless of column level); and (d) an error or residual. In verbal equation form this appears as:

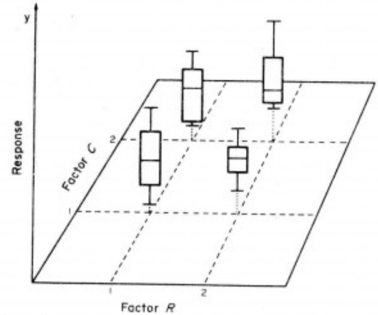

Figure 4
Graphical representation of hypothetical data for a two-way classification of responses (multiple unequal observations in each cell)

$$\text{Data} = \overbrace{\text{Common effect} + \text{Row effect} + \text{Column effect}}^{\text{Fit}} + \text{Residual} \qquad (6)$$

Whether the grand mean or grand median is used to estimate the common term, the result is conceptually identical. Its removal by subtraction effectively translates the origin plane so that the data are distributed around the origin rather than above it as in the hypothetical example. It is thus analogous to removing the grand median from a set of multiple batches, or the median from a single batch, and using the adjusted values, that is, the raw values with the grand median subtracted out, in constructing a new display.

By assuming that row and column effects are consistent, the model asserts that there will be only one effect for a row level regardless of the number of column levels, their size, and their effects and vice versa. In other words, the effect of row level 1 will be to elevate (or depress) the values in cells (1,1) and (1,2) the same amount, say a_1. Similarly, the effect of column level 2, according to the model, will be to elevate (or depress) the values in cells (1,2) and (2,2) the same amount, say b_2. Thus, any process of fitting this model to data must be constrained to finding these amounts according to a criterion that ensures an additive result. Given some estimate of these effects, the extent to which the data fail to conform can be studied by examining the residuals.

Thinking in terms of multiple observations in each