

## 中国武術の表演試合における問題点 (1)

—審判員間の信頼性について—

七 堂 利 幸<sup>\*</sup>(武当派(拳)研究会)

### I はじめに

わが国の中国武術の表演試合(型試合の意)は1983年から行われている。1987年に発足した日本武術太極拳連盟(以下、連盟と略す)が当該競技を主催している。

1990年七堂らは第1回、5回、6回当該大会の資料を分析して、中国武術の表演試合は採点競技として大きな問題をはらんでいることを本誌に報告した<sup>1)</sup>。それは、個々人に独立した採点方法という建て前とは別に、最初の選手を基準に採点し、順次前の選手に比べて優劣を見る相対的採点方法をとっているためである。後ろにどんな選手が出てくるか分からないという心理作用のため、得点は初めに抑えた点数をつけ漸次上昇していく典型的な上昇トレンドを示し、明らかに後半有利なエントリー順になっていた。その上、無作為に出場順位を決めていないことが原因で、表演試合以前にすでに不公平なエントリー順となっていた。

第11回北京アジア競技大会の予選である1990年3月の全日本大会から規定拳という統一套路の導入で競うようになった<sup>2)</sup>。

前報告<sup>1)</sup>の時点では、連盟は得点資料の公開拒否をしていたこともあって、採点の信頼性研究はできなかった。競技審判の信頼性には、審判員内と審判員間の二種の信頼性が考えられる。前者は同一審判員が場所を変え時間を変えた時、同一選手に再現性のある採点かどうかをみる。後者は同一選手を複数の審判員が採点してその一致度をみるものである<sup>3)</sup>。中国武術の表演試合は、5名の審判員の主観的な採点方法をとっている。本研究は、各審判員の採点データから審判員間の信頼性

について検討した。

### II 方法

材料は1994年の11回大会の予選と決勝のデータを使った。この11回大会は広島市で行われたアジア大会の予選でもあった。

#### (1) 規定拳の説明

規定拳とは、統一套路(套路は、中国語で型の意)で競うためのもので、楊式太極拳規定、陳式太極拳規定などという種目名で区分される。規定という名称がついていないが、他の統一套路(種目名)としては、24式太極拳、総合太極拳などがある。

1990年から始まった全日本大会から規定套路が大幅に取り入れられた。全日本大会は実績のある選手どうしの選抜大会で、第何回大会というのは過去から行われていた全国大会である。

規定套路で競う規定拳に対して、套路の内容が、競技者自作による独自のものは自選拳と呼ばれ、この種目には陳式太極拳、太極劍・刀、伝統器械などがある。器械とは、槍や刀などの武器のことである。

現在、表演大会はこれら規定套路と自選套路の二本立てで行われている。

(2) 表演大会の審判方法は、原則として、10点満点からの減点法(1/100の位まで表示)で、5名の各審判員(審判長1、審判員4名)の最高点、最低点を除いて、あとの3名の平均点で優劣を決める方法を取っている<sup>1)</sup>。

#### (3) 審判員間の信頼性

①手に入れる事が出来たデータをもとにして、表1のように第11回大会の36種目について、5名の審判員間の信頼性をピアソンの相関係数(以下rと略す)を用いて調べた。連関の強さはrで表わされ、-1.0と+1.0の間の値をとる。入手出来なかったデータは、女子総合、女子24式C、女

<sup>\*</sup> 〒543 大阪市天王寺区上本町6-3-31-1111号  
電話 06-779-0339

表1 審判間信頼性（1994年第11回表演大会）

審判比較	男子長拳規定		女子長拳規定		男子刀術器械		女子刀術器械		男子劍術規定		女子劍術規定		男子槍術規定		女子槍術規定		男子棍術規定		女子棍術規定		男子拳術規定		女子拳術規定		男子團體技		女子團體技	
	0.772	0.711	0.721	0.712	0.713	0.668	0.639	0.61	0.787	0.875	0.464																	
1:2	0.633	0.798	0.674	0	0.738	0.303	0.644	0.361	0.651	0.89	0.614																	
1:3	0.860	0.853	0.736	0.587	0.685	0.151	0.795	0.667	0.552	0.835	0.677																	
1:4	0.813	0.709	0.684	0.51	0.871	0.27	0.608	0.746	0.71	0.888	0.603																	
1:5	0.614	0.802	0.575	0.24	0.686	0.646	0.862	0.355	0.908	0.954	0.917																	
2:3	0.883	0.934	0.836	0.843	0.78	0.771	0.908	0.685	0.763	0.751	0.937																	
2:4	0.824	0.791	0.795	0.751	0.884	0.895	0.825	0.73	0.927	0.921	0.921																	
2:5	0.713	0.86	0.41	0.433	0.759	0.717	0.877	0.557	0.926	0.755	0.943																	
3:4	0.826	0.864	0.746	0.645	0.864	0.602	0.851	0.563	0.921	0.849	0.808																	
3:5	0.804	0.811	0.649	0.875	0.747	0.622	0.893	0.733	0.866	0.781	0.915																	
4:5	6	6	1	2	3	1	6	0	5	7	6																	
相闘数0.8以上の個数(個)	0.788	0.836	0.799	0.662	0.817	0.577	0.844	0.614	0.89	0.879	0.872																	
Kendallの一致係数W	0.01	0	0.01	0.06	0	0.24	0	0.13	0	0	0.01																	
独立性検定の有意確率(%)																												
上記欄目の決勝での審判間比較																												
1:2	0.841	0.941			0.939	0.758	0.817	0.84			0.837																	
1:3	0.842	0.908			0.696	0.851	0.798	0.445			0.801																	
1:4	0.874	0.957			0.861	0.561	0.837	0.737			0.927																	
1:5	0.946	0.93			0.871	0.792	0.901	0.823			0.725																	
2:3	0.702	0.88			0.667	0.763	0.963	0.556			0.945																	
2:4	0.739	0.91			0.807	0.85	0.908	0.839			0.65																	
2:5	0.845	0.893			0.836	0.742	0.955	0.959			0.751																	
3:4	0.805	0.854			0.931	0.438	0.925	0.753			0.535																	
3:5	0.860	0.837			0.753	0.885	0.92	0.573			0.767																	
4:5	0.942	0.834			0.834	0.413	0.902	0.927			0.571																	
相闘数0.8以上の個数(個)	8	10			7	3	8	5			4																	



七章：中国武術の表演試合における問題点（1）

表1の続き（2）

審判員比較	男子24式A	女子24式A	男子24式B	女子24式B	男子太極刀・劍	女子太極劍・刀	男子24式C	男子48式	男子総合本舞臺	男子形式判定	女子形式判定	女子形式規定
1:2	0.707	0.557	0.281	0.298	0.744	0.21	0.352	0.531	0.859	0.953	0.649	0.731
1:3	-0.083	0.387	0.604	-0.072	0.801	-0.156	0.326	0.329	0.756	0.604	0.116	0.071
1:4	0.338	0.573	0.396	0.323	0.788	0.699	0.759	0.588	0.859	0.481	0.628	0.665
1:5	0.561	0.482	0.895	-0.49	0.658	-0.194	0.345	0.065	0.742	0.904	0.558	0.673
2:3	-0.157	0.546	0.498	-0.148	0.897	0.135	0.801	0.234	0.906	0.576	0.056	0.13
2:4	0.597	0.595	0.502	0.79	0.898	0.34	0.506	0.547	0.91	0.612	0.736	0.765
2:5	0.497	0.427	0.331	0.17	0.725	0.445	0.704	0.414	0.725	0.919	0.682	0.958
3:4	0.534	-0.08	0.55	0.074	0.89	0.222	0.487	0.382	0.825	0.066	0.567	-0.381
3:5	0.644	-0.115	0.402	0.217	0.863	0	0.878	0.544	0.519	0.522	0.29	0.05
4:5	0.622	0.673	0.267	0.361	0.847	-0.099	0.27	0.546	0.792	0.551	0.579	0.752
相関係数0.8以上の個数(個)	0	0	1	0	6	0	2	0	5	3	0	1
Kendallの一致係数W	0.476	0.506	0.57	0.309	0.638	0.328	0.665	0.538	0.692	0.744	0.561	0.593
独立性検定の有意確率(%)	1.32	0.85	0.28	14.46	0.1	11.33	0.07	0.48	0.03	0.01	0.32	0.26
上記項目の決勝での審判員比較												
1:2	0.584	0.852	0.97	0.591					0.933			
1:3	0.489	0.508	0.884	0.578					0.854			
1:4	0.917	0.079	0.95	0.661					0.85			
1:5	0.457	0.41	0.343	0.603					0.982			
2:3	0.926	0.557	0.837	0.494					0.903			
2:4	0.731	0.039	0.952	0.597					0.892			
2:5	0.887	0.578	0.37	0.346					0.954			
3:4	0.651	0.016	0.949	0.528					0.939			
3:5	0.719	0.567	0.581	0.387					0.9			
4:5	0.648	-0.288	0.521	0.448					0.926			
相関係数0.8以上の個数(個)	3	1	6	0					10			

子48式の各太極拳，女子伝統拳術Aの4種目である。

また  $r$  はそのサンプル数により多少変動するので<sup>5)</sup>，ケンドールの一致係数（以下  $W$  とする）を求めその独立性検定を行った<sup>6)</sup>。この有意性は，理論的にフリーマン検定という一致性検定と同じことである<sup>7)</sup>。

一致係数  $W$  は， $r$  のように0.7以上が強い相関というようなコンセンサスはない。しかし， $r$  はスピアマンの順位相関係数と同一計算で求めることができ，後者と  $W$  は線型関係がある<sup>6)</sup>。そこで  $W$  に0.8という  $r$  と同じ基準で分類した。個々の審判員間を  $r$  で，種目単位では  $W$  で信頼性をみた。

予選順位10位までの選手について5名の審判員の採点を使った。女子南拳，女子棍術規定は選手が10名いなくて8名のデータで計算した。

②同一種目で予選と決勝の各審判員間の  $r$  を求め，信頼性を比較した。入手できた資料は12種目で， $r$  は120対120（個）でその平均と標準偏差を求め，0.8以上の  $r$  の個数も比較した。

### III 結果

体操競技が  $r$  の基準を0.8にしているので<sup>8)</sup>，それと比較するため，0.8という基準を設けた。

表1のように，全  $r$  の組み合わせは360組であった。全体として0.8以上の  $r$  はそのうち32%であった。全体を幹枝表示，箱型図（90%の信頼区間）にしたものが，図1である。中央値は0.708である。8個の  $r$  が負の外れ値表示されている。

全一致係数  $W$  は，中央値は0.695で，最低0.309，最高0.926であった。すべての  $W$  をみると，0.8以上は11/36=0.31という割合であった。

①最も信頼性の高いのは楊式太極拳規定で，男子は全審判員間の  $r$  が0.8以上で，女子は  $r$  が0.8以上の割合は9/10であった。 $W=0.907$ （男子）， $W=0.926$ （女子）であり，これは全審判員間の信頼性が高いことを示している。また，男子南拳規定で  $r$  が0.8以上の割合は9/10（ $W=0.893$ ），女子南拳規定は10/10（ $w=0.901$ ）で信頼性は高かった。最も低いのは24式太極拳で，全審判員間にも0.8以上の相関は無く， $r$  が最も低かった。 $W$  は男子を見ると，24式太極拳Aが0.476，24式太極拳Bが0.570，24式太極拳Cは0.665，女子は24式太極拳Aが0.506，24式太極拳Bは0.309であった。

②自選競技は自作の独自の套路で， $r$  が0.8以上の割合は表1のとおり，審判員間の信頼性は

	N	MEAN	MEDIAN	TRMEAN	STDEV	SEMEAN
C1	360	0.6247	0.7080	0.6492	0.2819	0.0149
	MIN	MAX	Q1	Q3		
C1	-0.4900	0.9770	0.4990	0.8445		

Stem-and-leaf of C1  
Leaf Unit = 0.010

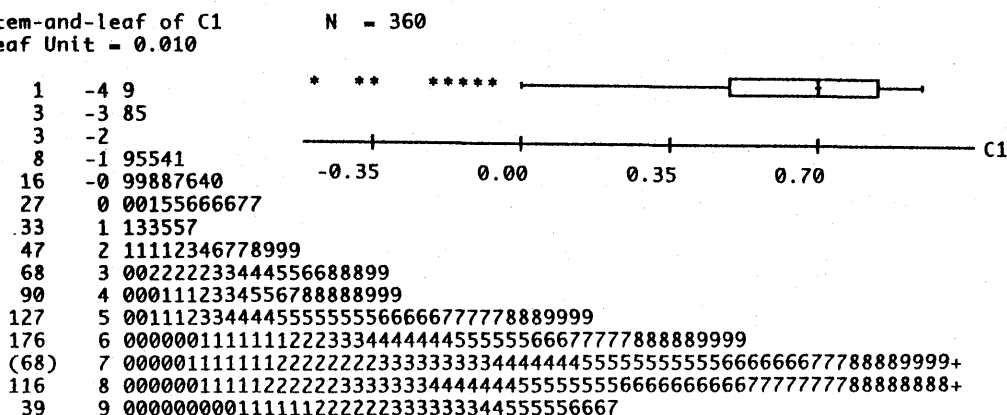


図1 相関係数  $r$  の記述統計値と幹枝表示，箱型図（90%信頼区間）

低い。

③規定競技は、 $r$  が 0.8 以上の割合は表 1 のとおりで、自選競技より信頼性は高いが、統一された動きで競っても、決して審判員間の信頼性は高くない。

④Wに関する独立性検定をみると、女子24式Bが有意水準  $p = 14.46$  (%)、女子楊式・88式が  $p = 6.76$  (%) で、女子太極剣・刀は  $p = 11.33$  (%) で順位づけに関連ないという帰無仮説は 5% 水準で棄却されない。つまり、審判員間の一致が偶然であることを表わす。逆にその他の種目 (33 / 36 種目) は、5人の審判の一致が偶然ではないことを表わす。Wの有意性検定が有意でもW値自身が大きくない。これは審判判定の一致度が確実に低いということである。

⑤12種目の予選と決勝の  $r$  を比較すると、表 1 のように  $r$  が 0.8 以上の数は、予選より (37個) 決勝 (72個) の方が多くなった。全  $r$  の平均値 ± 標準偏差を比較すると、予選は  $0.614 \pm 0.286$ 、決勝は  $0.748 \pm 0.226$  であった。その差は平均点で 0.134、 $r$  が 0.8 以上の数は  $72/37 = 1.95$  倍になった。以上の結果より、予選より決勝の方が、審判員間の信頼性が高かった。

#### IV 考察

1990年の報告<sup>1)</sup>で審判員間の信頼性に触れられなかったのは、当初連盟はデータの請求、公開を拒否してきたためである。そのため、審判員の質についての評価は困難であった。そこで、24式太極拳で予選と決勝の得点を利用して、同一審判と仮定した時の再試験法による信頼性に疑問があることを報告した<sup>1)</sup>。しかし今回は、最も知りたかった予選での各審判の採点結果が入手できたので、本研究が可能になった。

データの入手出来なかった4種目は、全体の48種目、解析対象36種目との比率より、結果に大きな変化をもたらす可能性は少ないと考えてよいだろう。

$r$  は、0.7以上という基準が、強い相関の一応のコンセンサスになっている<sup>7)</sup>。しかし、Johnson (1971) は、体操で  $r$  が 0.8以上という基準で信頼性係数として報告している<sup>8)</sup>。この研究は25年前の20年間の追跡調査で、体操コンパルソリー競

技が採点の変動が大きく、全体では  $r$  の 15% は 0.80 以下であったが、それでも結論は再評価の継続の必要性を強調している。中国武術である本研究結果において、0.8以上の  $r$  が全体で 32% しかない。これは 0.8 以下が 7 割弱もあるということになる。また全一致係数  $W$  も 0.8 以上は 31% で、あとの約 7 割が審判員間の一致が低い。 $r$  と  $W$  の% 一致の高さは、両者が線形関係にあるからである。

25年前の体操競技と比較すると、現在の中国武術の審判員間の信頼性は、0.8で分類した信頼性の高低の比率が逆転しており、非常に大きな問題があると言わざるを得ない。過去の1回大会から5回大会も、今回より優れた信頼性で審判していたとは考えられない。

1990年の前報告<sup>1)</sup>が本誌掲載後、直ちに同年から審判各人のごく一部の得点の公開が始まり、採点の公開により審判員の向上を目指したという。またこのときから、採点する審判以外に套路を規定どおりに演じているかどうかを見る套路検査員が置かれるようになった<sup>2)</sup>。しかし、本研究結果のように審判員間の信頼性は極めて低い。この低信頼性の原因は、採点基準のあいまいさや審判員の養成方法の問題であって、単に規定套路を導入したり、套路検査員をおくというような試合施行方法を変えれば改善できるというような末梢の問題ではないだろう。

本研究の継続調査<sup>10)</sup>では採点精度の改善があるとは言えないという結果になっており、「統一套路は印象点の要素が減り点数のばらつきが少なくなる」という連盟側の期待は<sup>2)</sup>はずれた形になっている。

①信頼性の最も低いのは本邦で最も普及している24式太極拳で、これは意外な結果であった。24式太極拳はすでに今で言う規定拳であったわけで、ここまでひどいとは予想できなかった。

選手数が減るために審判しやすい優勝戦でも、24式太極拳では男子で 0.8 以上の  $r$  の割合は 3 / 10、女子で 1 / 10 に改善しただけである。24式太極拳が本邦に導入されてすでに 30 年を超す。この一番慣れているはずの24式太極拳がどの審判員間にも 0.8 以上の相関は無く、 $r$  が計算した種目中最も低かった。

この24式太極拳が公認指導員の試験の中心になっている。また、最近導入した技能検定の段位制試験の主要種目でもある<sup>11)</sup>。特に審判員間信頼性のない種目でこのような資格試験を実施しているのは大きな問題である。

②男女とも楊式太極拳規定が最も審判員間の信頼性が高かった。これは規定拳だからといって、信頼性が高いわけではない。同じ楊式太極拳に内容の近い規定拳である男子総合太極拳では、 $r$ が0.8以上の割合は5/10、男子陳式では2/10である。ちなみに、この男子総合太極拳で選ばれた日本選手が中国選手の突然の怪我による棄権もあり、広島アジア大会では唯一金メダルを獲得した。

③一致係数 $W$ の大きな値や有意な値は、順序づけが正しいことを意味しない<sup>6)</sup>。つまり、高い一致は、客観性を保証するものではない。単に、同じ基準を適用していることを表わしているのである。審判基準の妥当性と、本論で問題にしている信頼性とは別の問題で、採点にはこの両者が要求される。 $W$ の有意水準が5%以下では5名の審判の順位が偶然ではないことを表わすし、 $W$ は10人の選手の順位の連関の度合を表わす。

$W$ とその独立性検定をみると、女子24式太極拳B、女子楊式・88式太極拳、女子太極剣・刀の3種目は、5名の審判が独立して採点していないことを表わしている。これらの3種目は $r$ も0.8以上あるものがないということで、審判員間の一致は見られず、対象種目中最悪の審判結果であったといえる。

全体の $W$ は、幹枝表示から0.7以上で1.0に近い数字が集まっているわけではなく、ピーク(峰)がなく分布はばらついて、0.5前後にも分布の塊があり、全体的に審判員間の信頼性は低い。

④信頼性の他に科学の要件として、妥当性の問題がある。ルールによって得点は異なる。中国武術の表演目的は「審美」といって見た眼の美しさを採点することになっている。採点が一致しない5名の審判員のうち一体誰が妥学な点数をつけているのか誰も分からないのである。 $W$ が高い値であっても、同じ基準を適用してその一致を見ているだけで、得点の妥当性は分からない。得点の妥当性はそのルールが本当に測定したいものを測定し

ているかという別の問題である。

⑤決勝の方が、 $r$ の平均値で0.134大きくなり、0.8以上の $r$ の個数では1.95倍増加している。予選より人数の限定される決勝の方が、明らかに審判員間の信頼性が高い。

これは前報告で予選より決勝の方が高得点で競われる現象と平行して起こっている。決勝が高得点で競われる現象は、第一に最初の選手の基準点が予選より高く設定されることがその理由である。急に選抜者全員が決勝でうまくなくなったわけではない。これは減点法を事実上とっていない証拠でもある。決勝で信頼性が高まるという本研究結果は、得点が予選より決勝が高得点で競われること<sup>10)</sup>や、少人数で競うことで、採点の幅が限定されるという審判員の必理からくると考えられる。

審判員間の信頼性がないので、信頼できない審判方法は妥当でないため、採点結果についてこれ以上論じる必要は無いかもしれない。しかし、審判方法の不備と選手エントリー順の公平性は全く別の問題である。選手エントリー順の公平性の問題を中心にして、次の報告(2)<sup>10)</sup>で検討したい。

## V まとめ

審判員間の信頼性について検討し、次のことが判明した。

①中国武術表演競技における審判員間の信頼性は、対象となった36種目においてピアソンの相関係数 $r$ が0.8以上の割合は全体で32%しかなく、ケンドールの全一致係数 $W$ も0.8以上が31%であった。

これは1971年の体操競技における審判員間の $r$ (15%は0.80以下)と比較すると、表演競技が審判員間の信頼性に大きな問題があることが分かった。

②規定套路は、自選套路に比較すると審判員間の信頼性数値はやや高いが、信頼性としては決して高いとは言えず、規定套路では審判方法の問題点が解決されていない。

③男女とも楊式太極拳規定と南拳規定が、最も審判員間の信頼性が高かった。最も低かったのは24式太極拳である。審判員間の信頼性から言うと、信頼性の低い24式太極拳( $W = 0.5$ )で昇級・昇段試験の技能検定や公認指導員試験を行うことは

## 七堂：中国武術の表演試合における問題点（1）

極めて問題であり、審判員間信頼性のあった楊式太極拳規定（ $W = 0.9$ ）でやり直すべきであろう。

④予選より人数の限定される決勝の方が、 $r$ の平均値で0.134大きくなり、0.8以上の $r$ の個数では1.95倍増加し、明らかに審判員間の信頼性が高くなった。これも異常なことである。

1990年に行われたというビデオを見ながら審判員が反省会を開いたという<sup>2)</sup>。反省会だけでは何の改善も得られないことは、以上の分析結果から明白である。

妥当とされない審判方法でも信頼性があるが、信頼性のない審判方法はすでに採点方法が妥当性を失っている。

統計ソフトはMINITAB,統計解析ハンドブックIV・ノンパラ編（共立出版）を使った。

### 文献

- 1) 七堂利幸, 朝倉宏之, 藤田誠：中国武術の型(表演)試合における審判方法の問題点, 武道学研究, 22(3), 57-64, 1990
- 2) 武術編集部：規定套路は表演武術シーンをどう変えるのか?, 季刊武術, 福昌堂, 72-73, 1990
- 3) Margaret J. Safrit (1980) : 体育アセスメントと評価, 遊佐・永田・宮崎・青山訳, 泰流社, 1982
- 4) 白旗慎吾：パソコン統計解析ハンドブックIX, 共立出版, 236-252, 336-343, 1987
- 5) J. D. Gibbons : Nonparametric measures of association, Sage . Publication, 32-40, 1993
- 6) S. ジーゲル：ノンパラメトリック統計学, 藤本監訳, マクロウヒル, 238-246, 1983
- 7) 森敏昭, 吉田寿夫編：心理学のためのデータ解析テクニカルブック, 北大路書房, 220, 1990
- 8) Marvin Johnson : Objectivity of judging at the national collegiate athletic association gymnastic meet : a twenty- year follow-up study, The Research, Quarterly, 44 (1), 454-455, 1971
- 9) 渡部洋：心理・教育のための多変量解析法入門, 福村出版, 11-12, 1992
- 10) 七堂利幸：中国武術の表演試合の問題点(2), - 5年後の追跡研究 -, 武道学研究29(1) : 34-42, 1996
- 11) 日本武術太極拳連盟編：太極拳の技能検定制度, 武術太極拳, 57, 6-8, 1994



## EXAMINATION OF JUDGING PROCEDURES FOR CHINESE MARTIAL ARTS FORM COMPETITIONS

### — Part 1. Research into the inter-rater reliability of judges —

**Toshiyuki SHICHIDO** (Martial Arts Research Group)

We found a major problem in the scoring procedures in our examination of the first, fifth, and sixth exhibitions. The judge use a system taking the first competitor as the standard against which all others are compared. This means that because they don't know what type of competitor will appear later on, there is psychological pressure which pushes scores upward as the competition progresses, producing a trend which favors competitors in the latter half.

Before considering whether or not there had been any improvement in the judging procedures for the Chinese martial arts form competitions over the last five years, we first looked at materials on the eleventh tournament (1994) to study the inter-rater reliability of judges. This covered 36 events. The results demonstrated the following ;

(1) This was the first examination of the inter-rater reliability of the judging for Chinese martial arts form competitions. The Pearson correlation coefficient  $r$  exceeding 0.8 was only 32% overall, and the Kendall coefficient of concordance  $W$  exceeding 0.8 was 31% overall.

On the other hand, the judging for the 1971 gymnastic competition is reported as showing 85%  $r$  at 0.80. Compared to this, there is obviously a big problem with the judging of Chinese martial arts.

(2) While the unified form judging demonstrates greater reliability than the individual form judging, the reliability is by no means high, and there are still problems remaining in the methods for unified judging as well.

(3) In speaking of inter-rater reliability of judging, we can say there are severe problems with the skills evaluations for advancement tests and testing of certified instructors for 24 form taichi rules, which show low reliability ( $W=0.5$ ). These should be re-evaluated using the Yang-style taichi rules, which demonstrate much high reliability ( $W=0.9$ ).

(4) On examining 12 events, the inter-rater reliability of judging for the championship round, involving limited numbers of competitors, was found to be clearly higher than for the preliminary rounds.

Even judging methods which are not considered adequate demonstrate reliability, but scoring procedures for unreliable judging methods are already clearly invalid.