

Aus dem Department der Internen Medizin Kleintiere  
der Veterinärmedizinischen Universität Wien

Universitätsklinik für Kleintiere  
(Leiter: Univ.-Prof. Dr.med.vet. Iwan Burgener Dipl.ECVIM-CA Dipl.ACVIM)

**Systematische Literaturübersicht zur  
Verhaltensbeurteilung von Einzelhunden im Tierheim und  
der Vergleich der Literatur**

Diplomarbeit

Veterinärmedizinische Universität Wien

vorgelegt von

Laura Wehner

Wien, im August 2021

Betreuerin: Dipl.ECAWBM-BM Dr<sup>in</sup> med.vet. Nadja Affenzeller

GutachterIn: Dipl.ECAWBM (AWSEL) Dr<sup>in</sup> med.vet. Christine Arhant

## Inhaltsverzeichnis

1	Einleitung und Fragestellung	1
1.1.	Der Haushund ( <i>Canis familiaris</i> ) im Tierheim	1
1.2.	Beurteilung von Hunden im Tierheim	2
1.3.	Ziel der Diplomarbeit: Erfüllen bereits publizierte Verhaltensevaluierungen die Anforderungen an diagnostische Testsysteme?	6
1.4.	Anforderungen an diagnostische Testsysteme	6
1.4.1.	Objektivität	7
1.4.2.	Reliabilität – Zuverlässigkeit	7
1.4.2.1.	Test-Retest-Reliabilität	8
1.4.2.2.	Inter-Rater-Reliabilität	8
1.4.2.3.	Intra-Rater-Reliabilität	8
1.4.2.4.	Interne Konsistenz	9
1.4.3.	Validität – diagnostische Gültigkeit	9
1.4.3.1.	Inhaltsvalidität	9
1.4.3.2.	Konstruktvalidität	9
1.4.3.2.1.	Konvergente Validität	9
1.4.3.2.2.	Diskriminante Validität	10
1.4.3.3.	Kriteriumsvalidität	10
1.4.3.3.1.	Übereinstimmungsvalidität	10
1.4.3.3.2.	Prognostische Validität	10
2	Material und Methoden	13
2.1.	Literatursuche	13
2.1.1.	Einschlusskriterien	13
2.1.2.	Ausschlusskriterien	13
2.1.3.	Informationsquellen	13
2.2.	Auswahl der Studien	15
2.2.1.	Identifikation	15
2.2.2.	Vorauswahl	18
2.2.3.	Eignung	20
2.2.4.	Einteilung in Evidenzklassen	29
3	Ergebnisse	31

3.1.	Einteilung der Studien in Evidenzklassen	31
3.2.	Verhaltensbeurteilung durch Fragebögen	35
3.2.1.	Fragebögen in der Praxis	35
3.2.1.1.	Verhaltensbeurteilung durch C-BARQ u. Modifikationen	35
3.2.1.2.	Verhaltensbeurteilung <i>nach STEPHEN et LEDGER 2007</i>	38
3.2.1.3.	Verhaltensbeurteilung durch MCPQ-R ( <i>“Monash“</i> )	39
3.2.2.	Vergleichende Übersicht aller Fragebögen	40
3.2.3.	Diagnostische Testkriterien der Fragebögen	43
3.2.3.1.	Objektivität von Fragebögen	43
3.2.3.2.	Reliabilität von Fragebögen	43
3.2.3.3.	Validität von Fragebögen	44
3.3.	Verhaltensbeurteilung durch Wesenstests	49
3.3.1.	Wesenstests in der Praxis	51
3.3.1.1.	Wesenstests in Europa	51
3.3.1.1.1.	Wesenstest <i>nach VAN DER BORG et al., 1991</i>	51
3.3.1.1.2.	<i>Ethotest©</i> und Modifikationen	52
3.3.1.1.3.	Wesenstest <i>nach FRANK, 2007</i>	54
3.3.1.1.4.	Wesenstest <i>nach Dogs Trust</i>	56
3.3.1.1.5.	Wesenstest <i>nach KLAUSZ et al., 2013</i>	57
3.3.1.1.6.	<i>Temperament-Test</i>	58
3.3.1.1.7.	Wesenstest <i>nach ZILOCCHI et al., 2015</i>	59
3.3.1.2.	Wesenstests in den USA	61
3.3.1.2.1.	<i>Assess-A-Pet</i>	61
3.3.1.2.2.	<i>SAFER™</i>	63
3.3.1.2.3.	<i>Evaluation of fearful dogs der ASPCA®</i>	66
3.3.1.2.4.	<i>Match-Up</i> und Modifikationen	67
3.3.1.3.	Wesenstests in Australien	69
3.3.1.3.1.	Wesenstest der <i>RSPCA</i>	69
3.3.1.3.2.	<i>B.A.R.K.</i>	70
3.3.2.	Vergleichende Übersicht aller Wesenstests	71
3.3.3.	Diagnostische Testkriterien der Wesenstests	77
3.3.3.1.	Objektivität von Wesenstests	77

3.3.3.2.	Reliabilität von Wesenstests	82
3.3.3.2.1.	Test-Retest-Reliabilität von Wesenstests	82
3.3.3.2.2.	Inter-Rater-Reliabilität von Wesenstests	89
3.3.3.2.3.	Intra-Rater-Reliabilität von Wesenstests	94
3.3.3.2.4.	Interne Konsistenz	97
3.3.3.3.	Validität von Wesenstests	98
3.3.3.3.1.	Inhaltsvalidität	98
3.3.3.3.2.	Kriteriumsvalidität/ prognostische Validität	98
3.4.	Verhaltensbeurteilung durch Ethogramme	112
3.4.1.	Ethogramme in der Praxis	112
3.4.1.1.	Ethogramm nach DE PALMA et al., 2005	112
3.4.1.2.	Ethogramm nach GOOLD et NEWBERRY, 2017	113
3.4.1.3.	Ethogramm nach CLAY et al., 2019	114
3.4.2.	Vergleichende Übersicht aller Ethogramme	114
3.4.3.	Diagnostische Testkriterien der Ethogramme	116
3.4.3.1.	Objektivität von Ethogrammen	116
3.4.3.2.	Reliabilität von Ethogrammen	116
3.4.3.3.	Validität von Ethogrammen	116
4	Diskussion	118
4.1.	Fragebögen	118
4.1.1.	Objektivität von Fragebögen	118
4.1.2.	Reliabilität von Fragebögen	120
4.1.2.1.	Test-Retest-Reliabilität von Fragebögen	120
4.1.2.2.	Inter-Rater-Reliabilität von Fragebögen	121
4.1.2.3.	Intra-Rater-Reliabilität von Fragebögen	122
4.1.3.	Validität von Fragebögen	123
4.1.4.	Conclusio Fragebögen	127
4.2.	Wesenstests	128
4.2.1.	Objektivität von Wesenstests	129
4.2.2.	Reliabilität von Wesenstests	132
4.2.2.1.	Test-Retest-Reliabilität von Wesenstests	133
4.2.2.2.	Inter-Rater-Reliabilität von Wesenstests	137

4.2.2.3.	Intra-Rater-Reliabilität von Wesenstests_____	140
4.2.3.	Prognostische Validität von Wesenstests_____	140
4.2.3.1.	Prognostische Validität ausgewählter Messparameter_____	143
4.2.3.1.1.	Aggression gegenüber Menschen_____	143
4.2.3.1.2.	Angst und Furcht_____	152
4.2.3.1.3.	Trennungsangst_____	153
4.2.3.1.4.	Freundlichkeit_____	155
4.2.3.1.5.	Verträglichkeit mit Artgenossen_____	155
4.2.3.1.6.	Umgänglichkeit mit Kindern_____	156
4.2.3.1.7.	Ressourcenverteidigung – Sonderfall Futter_____	157
4.2.3.1.8.	Vermittelbarkeit_____	160
4.2.4.	Einfluss der Abwesenheit einer Mensch-Hund-Beziehung_____	162
4.2.5.	Conclusio Wesenstests_____	165
4.3.	Ethogramme_____	172
4.3.1.	Objektivität von Ethogrammen_____	172
4.3.2.	Reliabilität von Ethogrammen_____	172
4.3.3.	Validität von Ethogrammen_____	173
4.3.4.	Conclusio Ethogramme_____	173
5	Zusammenfassung_____	175
6	Summary_____	177
7	Begriffserklärungen_____	179
8	Literaturverzeichnis_____	180
9	Tabellen-/Abbildungsverzeichnis_____	198

# 1 Einleitung und Fragestellung

## 1.1 Der Haushund (*Canis familiaris*) im Tierheim

In modernen Gesellschaften spielt der Haushund eine Vielzahl an wichtigen Rollen (MORNEMENT et al., 2009). In all diesen Rollen hängt der Erfolg des Hundes gravierend von dessen Verhalten ab – so muss ein Hund beispielsweise nicht nur die physischen Fähigkeiten zum Einsatz als Polizeihund, Suchhund oder Assistenzhund aufweisen, sondern auch charakterlich geeignet sein.

Es heißt, die Mensch-Hund-Beziehung ist die älteste Partnerschaft mit einem domestizierten Tier. Diese Beziehungen sind sehr komplex und werden möglicherweise problematisch, wenn sie nicht gut funktionieren. Aus einer nicht funktionellen Mensch-Hund-Beziehung können Verhaltensprobleme beim Hund resultieren. Unter Umständen werden diese Hunde eine Gefahr für alle Beteiligten und die Öffentlichkeit, indem sie Menschen oder andere Tiere verletzen (CANEJO-TEIXERA et al., 2018). In den USA werden beispielsweise über ein Prozent der Bevölkerung pro Jahr von Hunden gebissen und über 20 Menschen sterben pro Jahr an Hundebissen oder deren Folgen (HSU et SERPELL, 2003).

Verhaltensprobleme bei Hunden sind einer der Hauptgründe dafür, dass Hunde durch die BesitzerInnen vernachlässigt, ausgesetzt, euthanasiert, in Tierheimen abgegeben oder nach Adoption in eben diese zurückgegeben werden (DIESEL et al., 2008). Es wird geschätzt, dass zwischen 17 und 25 Prozent aller im Tierheim abgegeben Hunde aufgrund von Verhaltensproblemen abgegeben werden (HSU et SERPELL, 2003). Weiters können Verhaltensprobleme zu inadäquaten und/oder aversiven Trainingsmethoden von Seiten der BesitzerInnen gegenüber ihrer Tiere führen bzw. können inadäquate und/oder aversive Trainingsmethoden Verhaltensprobleme auslösen (FERNANDES et al., 2017). Daher sind Verhaltensprobleme und Trainingsmethoden mittlerweile eine Schlüsselfrage beim Thema Tierschutz. Die gängigsten verhaltensbedingten Gründe zur Abgabe im Tierheim sind Aggression gegen Menschen und/oder Tiere, wiederholtes Entlaufen, Zerstörung, Ungehorsam, Verunreinigung des Hauses und exzessives Bellen (SEGURSON et al., 2005). Neben diesen „hundebezogenen“ Gründen gibt es natürlich auch jene Gründe, welche auf die BesitzerInnen zurückzuführen sind, wie zum Beispiel Umzug, Zeitmangel oder Krankheit (MORNEMENT et al., 2010).

Weltweit gelangt eine Vielzahl von Hunden in Tierheime und Auffangstationen, allein im deutschsprachigen Raum landen geschätzt jährlich etwa 100.000 Hunde in Tierheimen (SÜDDEUTSCHE ZEITUNG, 2016; STATISTA RESEARCH DEPARTMENT, 2020; KOTRSCHAL et al., 2004). FRANK (2007) spricht in ihrer Arbeit sogar von 300.000 Hunden, die pro Jahr in deutschen Tierheimen landen. In den USA sind es jährlich sogar 3.9 Millionen Hunde (THIELKE et UDELL, 2019) in über 13.000 Tierheimen (GUNTER, 2008).

## 1.2 Beurteilung von Hunden im Tierheim

Tierheime stehen vor großen Herausforderungen, wenn es darum geht, Hunde mit Verhaltensproblemen zu versorgen und zu vermitteln (D'ARPINO et al. 2012).

Es ist sehr wichtig, dass die Verantwortlichen für die Vermittlung im Tierheim die Wesenseigenschaften der Hunde verlässlich einschätzen können, damit die Hunde in eine möglichst passende Umgebung vergeben werden und die neuen BesitzerInnen darauf aufmerksam gemacht werden können, was sie erwartet (DIESEL et al., 2008).

Aus diesem Grund gibt es sogenannte „*match*“-Programme, um die individuellen Bedürfnisse der Hunde und deren potenziellen Adoptivfamilien aufeinander abstimmen zu können. Ein Beispiel ist das „*Meet your Match Canine-ality™ adoption program*“ der US-amerikanischen Tierschutzorganisation *American Society for the Prevention of Cruelty to Animals (ASPCA®)*, 2021). Hierbei werden Erwartungen an den Hund via Fragebogen evaluiert (u.a. Informationen zu Hundeerfahrung, Verträglichkeit, Haltung, Beschäftigung, Alleinbleiben, Anhänglichkeit, Jagd-/Wachaufgaben, Charakter, speziellen Bedürfnissen).

Daher ist es notwendig, die Verhaltens- und Charaktereigenschaften der Hunde eruieren zu können. Es gibt verschiedene Möglichkeiten, das Verhalten und den Charakter von Hunden zu evaluieren: Wesenstests, Ethogramme, die Befragung der VorbesitzerInnen (falls möglich), die Befragung und Einschätzung durch das Tierheimpersonal und/oder ehrenamtlichen HelferInnen. Alternativ zu „Charakter“ werden in diesem Zusammenhang auch Bezeichnungen wie „Persönlichkeit“, „Temperament“ oder „emotionale Prädisposition“ verwendet, um die Eigenschaften eines Hundes zu beschreiben (TAYLOR et MILLS, 2006).

Weit verbreitet ist die Verwendung von Wesenstests in Tierheimen und anderen Organisationen, die Hunde in neue Zuhause vermitteln (SHABELANSKY et al., 2015). In



Europa werden beispielsweise in Zypern, Dänemark und Litauen Wesenstests bei Aufnahme im Tierheim durchgeführt (HAVERBEKE et al., 2015). Im englischsprachigen Raum sind die gängigsten Testverfahren zum Beispiel *SAFER™*, *Match-Up II* und *Assess-A-Pet* (ASPCA®, 2021).

In den USA haben in einer Umfrage 60.5% der befragten Tierschutzorganisationen angegeben, dass die Aufnahme von Hunden durch die verfügbaren Ressourcen, vor allem durch räumliche Kapazitäten, limitiert ist. Aufgrund dessen wenden 28.4% der befragten Organisationen formale Verhaltenstests an, davon machen 59.8% selbstentwickelte Tests aus (D'ARPINO et al., 2012). Auch in einer Umfrage in Australien gab ein Großteil der befragten Tierheime an, selbstentwickelte Verhaltensevaluierungen zu Hilfe zu nehmen, während die anderen Tierheime angaben, die bereits oben genannten Testverfahren anzuwenden, teilweise aber modifiziert (MORNEMENT et al., 2010).

Die Durchführung von Wesenstests ist bereits ein fixer Bestandteil in verschiedensten Bereichen. Neben der Verwendung in Tierheimen werden Begleit-, Assistenz-, Arbeits- und Sporthunde beurteilt (RAYMENT et al., 2015). Sogar ZüchterInnen verwenden derartige Tests, um ihre Welpen zukünftig an passende BesitzerInnen zu vergeben (MORNEMENT et al., 2009). Bei vielen Rassehunden ist die positive Absolvierung eines spezifischen Wesenstests Voraussetzung zur Zuchtauglichkeitszulassung. Auch der Zuchtausschluss besonders aggressiver Individuen kann ein Ziel von Wesenstests sein (NETTO et PLANTA, 1997). Das grundlegende Ziel von Wesenstests ist es daher, Hunde mit auffälligem Verhalten zu identifizieren und zu selektieren oder Hunde auf deren mögliche Ausbildung und/oder ihrer Einsatzfähigkeiten für bestimmte Tätigkeitsfelder zu überprüfen (BRADY et al., 2018).

Bei Wesenstests im Tierheim werden Hunde mit Stimuli konfrontiert, welchen sie voraussichtlich nach der Adoption begegnen werden (MATTERS, 2016). Das Ziel von Wesenstests im Tierheim soll es sein, diejenigen Hunde zu detektieren, welche bereit dazu sind, in die Vermittlung zu gehen und in weiterer Folge eine möglichst gute Übereinstimmung mit den passenden AdoptantInnen zu erreichen und/oder mögliche Verhaltensprobleme nach der Vermittlung vorauszusagen (CLAY et al., 2020a). Weiters sollen sie bei der Erstellung von erforderlichen individuellen Trainingsprogrammen helfen. Der durchgeführte Wesenstest stellt dabei jeweils eine Momentaufnahme der individuellen Verhaltenseigenschaften eines Hundes dar (MORNEMENT et al., 2015).

Aber auch zur Reevaluierung der Hunde, um etwaige Verhaltensänderungen nach einem längeren Aufenthalt im Tierheim zu beurteilen, soll der Wesenstest herangezogen werden können (FRANK, 2009). Verfügbare Ressourcen tragen jedoch maßgeblich dazu bei, wie weiter mit Hunden verfahren wird, welche den Wesenstest oder Teile davon nicht bestehen (MORNEMENT et al., 2010). So sind vor allem die finanziellen Mittel ausschlaggebend, aber auch die Verfügbarkeit von Zeit, Personal und Unterbringungsmöglichkeiten entscheiden darüber, ob diese Hunde weiterhin beherbergt und gegebenenfalls sogar aufwändigen Rehabilitationsprogrammen zugeführt werden können oder nicht.

Aus Gefährdungs- und Sicherheitsgründen wird ein besonderes Augenmerk auf die Detektion aggressiver Tendenzen bei einem Wesenstest gelegt (BOLLEN et HOROWITZ, 2008). Immer wieder geraten vor allem Wesenstests bei sogenannten „gelisteten“ oder „gefährlichen“ Hunden in die Schlagzeilen. So gibt es beispielsweise im deutschsprachigen Raum bundesland-/kantonabhängig eigene Vorschriften, welche einen Wesenstest bei manchen Hunden gesetzlich vorschreiben. So ist in Niedersachsen, Deutschland, die Haltung eines als gefährlich eingestuften Hundes unter anderen Auflagen auch an das Bestehen eines Wesenstests gebunden (THIESEN-MOUSSA et al., 2018).<sup>1</sup>

Auch Tierheime nehmen ihre Verantwortung zur Bewahrung der öffentlichen Sicherheit ernst und versuchen entsprechend zu handeln. (PATRONEK et BRADLEY, 2016). Hierbei soll vor allem eine irrtümliche Einstufung potenziell gefährlicher Hunde als vermittelbar vermieden werden (MORNEMENT et al., 2010). Die Verwendung des Begriffes „potenziell“ ist dadurch gerechtfertigt, dass das Auftauchen von Problemverhalten stark vom Umgang des zukünftigen Besitzers/ der zukünftigen Besitzerin abhängt (VAN DER BORG et al., 1991). Weiters vertritt jeder Mensch eine eigene Interpretation von „Problemverhalten“. Beispielsweise wird das Auftauchen von Futteraggression sehr stark beeinflusst vom Verhalten der AdoptantInnen (VAN DER BORG et al., 1991), so hat ein Viertel der NeubesitzerInnen einfach nicht mit dem Hund interagiert, während er gefressen hat und somit kein futteraggressives Verhalten ausgelöst. Auch unter jenen NeubesitzerInnen, welche Futteraggression wahrnahmen, wurde dies nicht als Problemverhalten gesehen. Futteraggressivität ist jedoch ein weit verbreiteter Einzeltest bei Wesenstests.

---

<sup>1</sup> gemäß dem niedersächsischen Gesetz über das Halten von Hunden gilt ein Hund als ein gefährlicher Hund, wenn Tatsachen den Verdacht rechtfertigen, dass von dem Hund eine Gefahr für die öffentliche Sicherheit ausgeht.

Bei manchen Hunden ergibt der Wesenstest, dass sie weder zur Adoption, noch zu einem weiteren Training geeignet sind. In manchen Ländern führt dies zur Euthanasie des jeweiligen Tieres. So werden Hunde, die durch Wesenstests fallen, in Australien in weiterer Folge euthanasiert (MORNEMENT et al., 2014). Die Art des Vorgehens ist innerhalb der EU sehr uneinheitlich. In Dänemark dürfen „gefährliche Hunde“ nicht adoptiert werden (HAVERBEKE et al., 2015). In Italien und anderen Ländern der EU dürfen Hunde in Tierheimen nur aufgrund schwerer oder unbehandelbarer Krankheiten oder aufgrund von Gefährlichkeit eingeschläfert werden (ZILLOCCHI et al., 2015).

Konkret in Österreich dürfen Hunde im Tierheim prinzipiell nur mit demselben „vernünftigen Grund“ euthanasiert werden wie Hunde in privater Obhut (BINDER, 2018). Somit ist eine Tötung eines Hundes im Tierheim nur im Einzelfall gerechtfertigt. Dies ist der Fall, wenn der Hund unter einer nicht behandelbaren Krankheit leidet oder dem Hund aufgrund von geringer Vermittlungschancen und den damit verbundenen institutionellen Haltungsbedingungen langfristig Leid zugefügt werden würde.

Aus der sogenannten „no-kill-Politik“ einiger Staaten, besonders in Europa, ergibt sich die Tatsache, dass einige Hunde viele Jahre im Tierheim verbringen (BARNARD et al., 2016). Ohne die unbestreitbaren Vorteile dieser Verfahrensweise zu vernachlässigen, hat sie dennoch eine chronische Überpopulation in die Tierheime gebracht und somit höhere Kosten zu deren Erhaltung. Es stellt sich aber auch die Frage nach der Lebensqualität besonders derer Hunde, die lebenslang im Tierheim leben (CAFAZZO et al., 2014). Das Problem der überfüllten Tierheime ergibt sich unweigerlich aus niedrigeren Adoptionszahlen als Eingangszahlen (DALLA VILLA, 2013).

Obwohl Wesenstests zahlreich und seit Jahrzehnten in vielen verschiedenen Kontexten angewendet werden, geraten sie immer wieder in die Kritik, da es ihnen an Einheitlichkeit und wissenschaftlicher Validierung fehlt (MORNEMENT et al., 2014). Da diese Tests jedoch vor allem in Tierheimen oft über Leben und Tod entscheiden, ergibt sich daraus ein Tierschutzanliegen - die Konsequenzen für den einzelnen Hund, welcher als verhaltensbedingt unvermittelbar identifiziert wurde, können düster sein (RAYMENT et al., 2015).

### **1.3 Ziel der Diplomarbeit: Erfüllen bereits publizierte Verhaltensevaluierungen die Anforderungen an diagnostische Testsysteme?**

Es existieren bereits vielfältige Verfahren zur Beurteilung der Vermittelbarkeit, des Aggressionspotentials und/oder des Charakters von Einzelhunden im Tierheim.

Verhaltensbeurteilungen von Tierheimtieren sollen als standardisierte Versuchssituationen definiert werden, in denen Testsituationen und Stimuli dazu dienen, ein Verhalten hervorzurufen, zu dokumentieren und statistisch mit dem anderer Individuen in der gleichen Situation zu vergleichen. Das Ziel ist es, das getestete Tier zu klassifizieren (JONES et GOSLING, 2005).

Der Zweck eines Tests muss vorsichtig abgewogen werden (TAYLOR et MILLS, 2006). Hierbei spielen die Messparameter eine entscheidende Rolle. Welche Messparameter sollen gemessen werden? Welche Verhaltensweisen sollen durch den Test aufgedeckt werden? Der Zweck ist der Schlüssel zur Methode und somit zu dem Aufbau der Einzeltests, bzw. der Einzelfragen. Beispiele hierfür sind die Identifikation von Aggression oder Angst, im Allgemeinen oder gegenüber einem spezifischen Reiz. Diese Zwecke variieren zwischen den diversen Evaluierungsverfahren, abhängig davon, welche Priorität die Selektierenden setzen.

Ich möchte im Zuge der Diplomarbeit auf die Praxistauglichkeit und Verlässlichkeit dieser bereits beschriebenen Methoden und deren Inhalte eingehen und vorhandene Veröffentlichungen auf die Erfüllung wichtiger diagnostischer Testparameter untersuchen. In einem ersten Schritt wird daher die Qualität der bereits publizierten Studien beurteilt und diese in Evidenzklassen eingeteilt. Anschließend werden die Studienergebnisse auf die Angaben von relevanten diagnostischen Testparametern untersucht und verglichen.

Die Hypothese dieser Arbeit lautet, dass keines dieser publizierten Verfahren ausreichend standardisiert bzw. wissenschaftlich validiert ist, um die Anforderungen an ein diagnostisches Testverfahren zu erfüllen.

### **1.4 Anforderungen an diagnostische Testsysteme**

Im Folgenden werden die wichtigsten Kriterien beschrieben, die bei der Bewertung der Güte von Resultaten eines Verfahrens angewandt werden können. Leider wurde in den Studien nicht einheitlich vorgegangen, da die AutorInnen ihren Fokus sehr unterschiedlich setzten.

### 1.4.1 Objektivität

Die Objektivität misst die Unabhängigkeit der Ergebnisse von dem Beobachter/ der Beobachterin (ETTLIN et GALLO, 2019). Diese/r sollte sowohl die Datenerhebung, als auch die Datenauswertung und die Dateninterpretation nicht beeinflussen können.

Die Abgrenzung zwischen der Beurteilung der Objektivität und der Reliabilität ist nicht ganz eindeutig (die Zuverlässigkeit setzt Objektivität voraus).

Weiters zählt das Vorhandensein eines standardisierten Ablaufs zur Bewertung der Objektivität (BUSER et al., 2007). Sind beispielsweise im Wesenstest die Testinstruktionen schriftlich und exakt festgehalten? Sind der Testort, die Testpersonen, die Stimuli, die Länge der Einzeltests und der Testzeitpunkt standardisiert? Ziel ist es, sämtliche potentielle Quellen für Variabilität zwischen den Tests zu identifizieren und zu kontrollieren, damit die einzige Variable die Reaktion des Hundes ist (TAYLOR et MILLS, 2006). Die Standardisierung des Tests hilft dabei, die Test-Retest-Reliabilität zu steigern.

Wurde auf den Arbeitsblättern im Wesenstest das Verhalten genau beschrieben (z.B. *Knurren, Erstarren, Rute eingezogen,...*) oder wurde direkt interpretiert (z.B. *ängstlich, aggressiv,...*)? Wurde der Wesenstest von einer Person durchgeführt und direkt beurteilt oder gab es eine klare Abgrenzung zwischen durchführender Person und beurteilender Person?

Auch die Praktikabilität eines Tests spielt eine Rolle dabei, ob er großflächig verwendet werden kann (TAYLOR et MILLS, 2006). Unpraktische, übermäßig lange oder komplizierte Tests werden mit hoher Wahrscheinlichkeit nicht exakt und zuverlässig umgesetzt werden können. Da die Standardisierung von Beurteilungsverfahren entscheidend für die Praktikabilität ist, ordne ich sie hier der Objektivität zu, ebenso wie die Frage, wie leicht eine Entscheidung über den Hund, bezugnehmend auf Testergebnisse, getroffen werden kann.

### 1.4.2 Reliabilität – Zuverlässigkeit

Die Reliabilität misst die Zuverlässigkeit der erhobenen Daten. Das bedeutet, sind die Testergebnisse unter gleichen Bedingungen reproduzierbar? Liefert der Test auch bei wiederholter Messung dieselben Ergebnisse? (BUSER et al., 2007).

Die Reliabilität behandelt den Grad, zu dem ein Test frei von Messfehlern ist (TAYLOR et MILLS, 2006). Die Reliabilität ergibt sich aus der Konsistenz der Resultate in den Einzel- und Gesamttests, zwischen verschiedenen BeurteilerInnen, in verschiedenen Umgebungen etc. Die Reliabilität ist die Voraussetzung für die Validität eines Tests. Die Reliabilität setzt der Validität eine Obergrenze (PATRONEK et al., 2019). Bei einer Reliabilität von unter 100% muss der Umfang etwaiger Ansprüche an die Validität entsprechend reduziert werden.

BRADY et al. (2018) nehmen zur Beurteilung der Zuverlässigkeit (*reliability*) folgende Parameter zur Hand:

- Inter-Rater-Reliabilität (*inter-rater reliability*)
- Intra-Rater-Reliabilität (*within-session intra-rater reliability*)
- Wiederholpräzision/ Test-Retest-Reliabilität (*test-retest reliability*)

#### **1.4.2.1 Test-Retest-Reliabilität**

Die Test-Retest-Reliabilität beschreibt das Ausmaß, in dem Verhaltenstests Eigenschaften eines Hundes unabhängig von Zeit und Kontext identifizieren können. Die Ergebnisse des Individuums sollten konsistent, sprich verallgemeinerbar sein zu verschiedenen Zeitpunkten und Gegebenheiten (BRADY et al., 2018). Man kann die Wiederholpräzision als Korrelationskoeffizient  $r$  zwischen den beiden Messungen ausdrücken (BUSER et al., 2007).

#### **1.4.2.2 Inter-Rater-Reliabilität**

Die Inter-Rater-Reliabilität beschreibt das Ausmaß, wie unterschiedlich verschiedene BeobachterInnen dasselbe Individuum auf dieselbe Weise beschreiben. Erzielen zwei unterschiedliche und voneinander unabhängige Personen die gleichen Testergebnisse (BRADY et al., 2018)? PATRONEK et al. (2019) bezeichnen die Inter-Rater-Reliabilität als den wichtigsten Typ der Reliabilität.

#### **1.4.2.3 Intra-Rater-Reliabilität**

Die Intra-Rater-Reliabilität beschreibt das Ausmaß der Übereinstimmung, die ein/e BeobachterIn zu verschiedenen Zeitpunkten hinsichtlich eines konstant beobachteten Kriteriums macht. Erzielt eine Person dieselben Testergebnisse, wenn sie diese mehrfach und ident ablaufend durchführt (BRADY et al., 2018)?

#### **1.4.2.4 Interne Konsistenz**

Die interne Konsistenz (auch *internal reliability* oder *internal consistency*) misst die Homogenität einer Erhebung (ETTLIN et GALLO, 2019) und ist somit ebenfalls eine Methode zur Abschätzung der Reliabilität. Sie bewertet, inwiefern Einzelfragen eines Fragebogens oder Einzeltests eines Testverfahrens den gleichen diagnostischen Aspekt messen und miteinander in Beziehung stehen. Diese Übereinstimmung wird in der Maßzahl *Cronbach's Alpha* angegeben (BUSER et al., 2007). Je höher die interne Konsistenz, gemessen im Kennwert *Cronbachs  $\alpha$* , desto höher ist die Reliabilität der Untersuchung. Werte über oder gleich 0.7 gelten als akzeptabel (WIENER et HASKELL, 2016).

#### **1.4.3 Validität - diagnostische Gültigkeit**

Die Validität erfasst die Eignung, die Aussagekraft und Brauchbarkeit der spezifischen Schlussfolgerungen, welche aus den Testergebnissen gezogen werden (TAYLOR et MILLS, 2006). Im Allgemeinen kann man sagen, dass eine Messung valide ist, wenn sie misst, was sie messen soll.

##### **1.4.3.1 Inhaltsvalidität**

Die Inhaltsvalidität (*content validity*) gibt an, ob der Test misst, was er messen soll und ob die Messkomponenten alle Aspekte der Messparameter abdecken (TAYLOR et MILLS, 2006). Die Inhaltsvalidität ist gegeben, wenn das gewählte Messverfahren bestmöglich die Messparameter misst. Ein Aspekt der Inhaltsvalidität ist die Anschauungs-Validität (*face validity*). Sie gibt an, wie sehr ein Test für Laien plausibel erscheint, sagt aber weiterhin nichts über die wahre Validität aus.

##### **1.4.3.2 Konstruktvalidität**

Die Konstrukt-Validität (*construct validity*) gibt an, in welchem Rahmen die Einzeltests bzw. Einzelfragen den Messparameter messen, den sie messen sollen (TAYLOR et MILLS, 2006). Zur Konstrukt-Validität zählen die konvergente und die diskriminante Validität.

###### **1.4.3.2.1 Konvergente Validität**

Konvergente Validität (*convergent validity*) liegt vor, wenn theoretisch miteinander in Verbindung stehende Faktoren tatsächlich miteinander korrelieren (TAYLOR et MILLS,

2006). Sie gibt an, wie sehr Tests für zusammenhängende Konstrukte miteinander korrelieren. Beispielsweise erwarten wir von einem Test zur Identifikation von Aggression bei Hunden, dass positiv Getestete eine höhere Euthanasierate haben als negativ Getestete.

#### **1.4.3.2 Diskriminante Validität**

Diskriminante Validität (*discriminant validity*) liegt vor, wenn theoretisch nicht miteinander in Verbindung stehende Faktoren tatsächlich nicht korrelieren (TAYLOR et MILLS, 2006).

#### **1.4.3.3 Kriteriumsvalidität**

Die Kriteriumsvalidität (*criterion validity*) gibt an, wie stark der Zusammenhang zwischen den Messergebnissen und einem externen Kriteriums ist (TAYLOR et MILLS, 2006). Zur Kriteriumsvalidität zählen die Übereinstimmungsvalidität und die prognostische Validität.

##### **1.4.3.3.1 Übereinstimmungsvalidität**

Die Übereinstimmungsvalidität (*concurrent validity*) kann gemessen werden, wenn es bereits ein validen Test (als externes Kriterium) gibt, welcher mit dem zu validierenden Test verglichen werden kann (TAYLOR et MILLS, 2006). Diese Form der Kriteriumsvalidität spielt in dieser Arbeit keine Rolle.

##### **1.4.3.3.2 Prognostische Validität**

Die prognostische Validität (*predictive value*) wird gemessen, wenn der zu validierende Test zeitlich vor der Erhebung des externen Kriteriums liegt (TAYLOR et MILLS, 2006). Das Ziel eines diagnostischen Tests ist im Falle von Verhaltensbeurteilungen im Tierheim immer ein möglichst hoher Vorhersagewert. Lassen sich basierend auf den Testergebnissen Aussagen über die Zukunft ableiten?

Die Validität (*validity*) eines Tests wird bestimmt durch dessen Sensitivität und Spezifität (PATRONEK et BRADLEY, 2016). Durch eine möglichst hohe Sensitivität, Spezifität und hoher Vorhersagewerte wird ein Test genau, die Trennschärfe zwischen „positiv“ und „negativ“ (oder anderen binären Klassifikatoren) ist hoch. Mit anderen Worten: misst ein Test tatsächlich das, was er messen soll?



Die Sensitivität ist der Prozentsatz von Individuen, welche tatsächlich positiv sind für den Zustand, den der Test als positiv identifiziert hat, sprich „Wie viele tatsächlich Positive werden durch den Test als positiv erkannt?“. Sie berechnet sich aus  $a / (a + b)$ , siehe Tab. 1.

Die Spezifität ist der Prozentsatz von Individuen, welche tatsächlich negativ sind für den Zustand, den der Test als negativ identifiziert hat, sprich „Wie viele tatsächlich Negative werden durch den Test als negativ erkannt?“. Sie berechnet sich aus  $d / (c + d)$ , siehe Tab. 1.

Spezifität und Sensitivität haben direkten Einfluss auf die Fähigkeit, ein gültiges Ergebnis vorauszusagen, sind aber in Verhaltensfragen sehr schwierig ermittelbar. Eine Schlüsselkomponente zur Berechnung von Spezifität und Sensitivität ist ein Referenzwert. PATRONEK et BRADLEY nahmen hierfür den Wert „Reaktion des Hundes nach der Adoption“. Dies ist notwendig, um die Rate Falsch-Positiver und Falsch-Negativer zu berechnen, ist aber aus ethischen und praktischen Gründen nicht immer komplett umsetzbar.

Der positive prädiktive Wert, auch „Prädiktiver Wert eines positiven Tests“, gibt an, wie viele positiv getestete Individuen auch tatsächlich positiv sind. Er berechnet sich aus  $a / (a + c)$ , siehe Tab. 1. Beispielsweise gibt er an, ob ein Hund, der im Wesenstest positiv auf das Vorhandensein von Verhaltensproblemen getestet wurde, auch tatsächlich später Verhaltensprobleme aufweist. Im Unterschied dazu, sagt die Sensitivität aus, ob ein Hund mit Verhaltensproblemen im Test auch tatsächlich positiv getestet wird.

Der negative prädiktive Wert, auch „Prädiktiver Wert eines negativen Tests“, gibt an, wie viele negativ getestete Individuen auch tatsächlich negativ sind, später also keine Verhaltensprobleme zeigen. Er berechnet sich aus  $d / (b + d)$ , siehe Tab. 1. Im Unterschied dazu sagt die Spezifität aus, ob ein Hund ohne Verhaltensprobleme auch tatsächlich negativ getestet wird.

Die Prävalenz verbindet die Gültigkeit eines Tests (Spezifität und Sensitivität) mit den Gegebenheiten in der realen Welt. Sie gibt Hinweis auf die Häufigkeit einer Krankheit, eines Symptoms oder in unserem Falle eines Verhaltensproblems in einer Population zu einem bestimmten Zeitpunkt. Bei einer niedrigen Prävalenz ist der negative prädiktive Wert hoch (es gibt also wenige Falsch-Negative) und der positive prädiktive Wert ist niedrig (es gibt also viele Falsch-Positive).

Interessant ist außerdem die Vortestwahrscheinlichkeit, welche sich auf eine Selektion der Population bezieht (ETTLIN et GALLO, 2019). In unserem Falle wäre das die Wahrscheinlichkeit dafür, dass ein bestimmtes (Problem-)Verhalten in der Population von abgegebenen Hunden im Tierheim auftritt. Diese weicht vermutlich von der Prävalenz ab, da Verhaltensprobleme ein Hauptgrund für die Abgabe im Tierheim ist und daher die Vortestwahrscheinlichkeit größer ist als die Prävalenz in der Durchschnittspopulation.

Tab. 1: Tabelle nach PATRONEK et BRADLEY, 2016

<b>Ergebnis der Verhaltensbeurteilung im Tierheim</b>	<b>Verhalten des Hundes nach der Adoption</b>	
	<b>Hund zeigt Verhaltensprobleme</b>	<b>Hund zeigt keine Verhaltensprobleme</b>
<b>Testergebnis positiv</b> (Problemverhalten wurde detektiert)	<b>tatsächlich positive Hunde (a)</b>  (Testergebnis positiv UND Hund wird Verhaltensprobleme zeigen)	<b>Falsch-positive Hunde (c)</b>  (Testergebnis positiv, aber der Hund wird kein Problemverhalten zeigen)
<b>Testergebnis negativ</b> (kein Problemverhalten detektiert)	<b>Falsch-negative Hunde (b)</b>  (Testergebnis negativ, aber der Hund wird Problemverhalten zeigen)	<b>tatsächlich negative Hunde (d)</b>  (Testergebnis negativ UND der Hund wird kein Problemverhalten zeigen)

## **2 Material und Methoden**

Übersichtsarbeiten erfordern ein gewisses Maß an Klarheit und Qualität (MOHER et al., 2009). Im Jahre 2005 wurde in Ottawa, Kanada, das PRISMA Statement entwickelt mit dem Ziel das Verfassen von systematischen Übersichten und Meta-Analysen zu verbessern. PRISMA steht für "Preferred Reporting Items for Systematic Review and Meta-Analyses", die deutsche Übersetzung lautet: „Bevorzugte Report Items für systematische Übersichten und Meta-Analysen“. Die Vorgehensweise beinhaltet eine 27-teilige Checkliste und ein Flussdiagramm, aufgeteilt in vier Phasen (PRISMA, 2020). Auch bei der kritischen Bewertung bereits publizierter Übersichtsarbeiten kann PRISMA hilfreich sein.

### **2.1 Literatursuche**

#### **2.1.1 Einschlusskriterien**

Bezüglich der Merkmale der Studien wurde zunächst großzügig jede Studie einbegriffen, die sich mit der Beurteilung von Verhalten der Hunde im Tierheim befasst. Es wurden Studien in deutscher und englischer Sprache gesucht. Es wurden nur Studien berücksichtigt, die online über den Hochschulzugang der Veterinärmedizinischen Universität Wien zur Gänze einsehbar waren.

#### **2.1.2 Ausschlusskriterien**

Einige Studien mit englischem Titel und Abstract haben sich bei der Beurteilung des Volltextes als Studien in spanischer und portugiesischer Sprache herausgestellt und wurden exkludiert. Keine der Studien wurde vor mehr als 29 Jahren publiziert, allerdings wurden keine älteren Studien zum Thema gefunden, da vermutet wird, dass das Thema in den letzten Jahrzehnten erst an Bedeutung zugenommen hat.

#### **2.1.3 Informationsquellen**

Die Studien wurden erstmals im September 2019 in verschiedenen Datenbanken gesucht (z.B. PubMed®, siehe Tab. 2). In weiterer Folge wurde in den Quellenverzeichnissen bereits gefundener Studien nach weiteren Veröffentlichungen zum Thema gesucht. WissenschaftlerInnen, welche sich laut vorhandener Studien offensichtlich im Zuge ihrer Laufbahn mit dem Thema befassten, wurden in der ForscherInnen-Datenbank *Research-Gate* nachgeschlagen

und nach weiteren themenrelevanten Publikationen durchsucht. Des Weiteren ergaben die Informationsbroschüren zweier etablierter Beurteilungsverfahren jeweils Anmerkungen zu deren Forschungsgrundlagen. Diese wären das Buch *“Assessing Aggression Thresholds in Dogs“* (Sue Sternberg, 2017) und *SAFER™* (ASPCA®, 2007).

Die Suche mit den vordefinierten Suchbegriffen (siehe Tab. 2) wurde am 31. Oktober 2020 abgeschlossen.

Tab. 2: Kriterien und Suchbegriffe der Literaturrecherche zu Wesenstests von Hunden im Tierheim

Parameter	Kriterien
Sprache	Englisch und Deutsch
Zeitraum	Beginn: September 2019 Ende: Oktober 2020
Suchmaschinen	Pubmed® (National Library of Medicine) Katalog der Vetmeduni Vienna Google/ Google Scholar Mendeley (2020 Elsevier B.V.) Science direct (2020 Elsevier B.V.) Scopus (2020 Elsevier B.V.) Web of Science (Clarivate Analytics)
Suchbegriffe	behavior assessment + dog + shelter behavior evaluation + dog + shelter shelter assessment + dog temperament test + dog + shelter Match Up behavior evaluation + dog + shelter SAFER + behavior + dog + shelter Assess-A-Pet + shelter C-BARQ + shelter aggression evaluation + dog + shelter aggression assessment + dog + shelter Wesensbeurteilung + Tierheim + Hund Verhaltensbeurteilung + Tierheim + Hund Wesenstest + Tierheim + Hund Beurteilung + Tierheim + Hund
Weiters	Suche im Literaturverzeichnis gefundener Studien ForscherInnen Datenbank ResearchGate zugrundeliegende Studien etablierter Beurteilungsverfahren

## 2.2 Auswahl der Studien

### 2.2.1 Identifikation

Folgende Schlüsselwörter wurden bei der Suche verwendet: *behavior assessment, behavior evaluation, shelter assessment, temperament test, Match Up behavior evaluation, SAFER™, Assess-A-pet, C-BARQ, aggression evaluation, aggression assessment, Wesensbeurteilung, Verhaltensbeurteilung, Wesenstest* (siehe Tab. 2).

Da eine Suche nach *“behavior assessment“* in PubMed® beispielsweise 734.088 Treffer ergab, war ein erster Reduktionsschritt notwendig. Zur Eingrenzung wurde jeweils *“dog“*, *„Hund“* und *“shelter“*, *„Tierheim“* ins Suchfeld hinzugefügt, um möglichst präzise Suchergebnisse zu erhalten.

Die Suche der Schlüsselwörter *“behavior assessment“* in Kombination mit *“dog“* und *“shelter“* erbrachte in PubMed® beispielsweise nur mehr 76 Treffer. Identifiziert wurden sämtliche wissenschaftliche Beiträge, welche das Thema entfernt angeschnitten haben und scheinbaren Bezug zur Fragestellung aufwiesen. So wurden alle 76 Treffer anhand ihres Titels in die Identifikation eingeschlossen oder ausgeschlossen. Eingeschlossen wurden sämtliche wissenschaftliche Beiträge, welche sich mit Beurteilungsverfahren von Hunden im Tierheim befassen.

Ausgeschlossen wurden Studien, welche sich zwar durch die Suchbegriffe in der Trefferliste befanden, allerdings bereits anhand ihres Titels nicht zum Thema passen, da sie:

- sich zwar mit dem Verhalten von Hunden in Tierheimen beschäftigen, allerdings keinen Hinweis auf entsprechende Beurteilungsverfahren geben

Bsp.: *“In-kennel behavior predicts length of stay in shelter dogs.“* (PROTOPOPOVA et al., 2014).

Die Situation im Tierheim ist eine außergewöhnliche Situation für den zu testenden Hund aufgrund der überfordernden Umgebung, die das Zeigen der Persönlichkeit (welche der Hund in einer vertrauten Umgebung zeigen würde) hemmen könnte (VALSECCHI et al., 2011). Daher ist die Beurteilung von Hunden im Tierheim nicht gleichzusetzen mit der Beurteilung von Hunden in anderen Situationen. Studien, welche Verhaltensevaluierungen von Militärhunden, Assistenzhunden, Rassen zur Zuchtauswahl u.a. untersuchen, wurden

daher ebenso exkludiert sofern kein Tierheimbezug besteht. Somit wurden jene Studien ausgeschlossen Studien, welche:

- sich mit Verhaltensproblemen von Hunden bei BesitzerInnen befassen und somit weder Bezug zu deren Evaluierung noch zur Tierheimsituation aufweisen.

Bsp.: *“Pet owners’ views of pet behavior problems and willingness to consult experts for assistance.”* (SHORE et al., 2008)

- sich mit der Verhaltensforschung bestimmter Hunderassen/-typen befassen, allerdings ebenfalls keinen Hinweis auf Evaluierungsverfahren geben.

Bsp.: *“Relationship Between Scarring and Dog Aggression in Pit Bull-Type Dogs Involved in Organized Dogfighting”* (MILLER et al., 2016)

Bsp.: *“Behavior Differences Between Search-and-Rescue and Pet Dogs”* (HARE et al., 2018)

- sich mit Optimierungs- oder Modifikationsmöglichkeiten von Tierheimhaltung oder mit Therapie- oder Trainingsmöglichkeiten von Hunden im Tierheim befassen und somit nicht relevant für die Wesensbeurteilung der Hunde an sich sind.

Bsp.: *“The effects of exercise and calm interactions on in-kennel behavior of shelter dogs.”* (PROTOPOPOVA et al., 2018)

- sich mit der Beurteilung und/oder Befragung von TierärztInnen, TierheimmitarbeiterInnen und/oder AdoptantInnen befassen und somit nicht mit der Beurteilung der Hunde.

Bsp.: *“Veterinary Behavior: Assessment of Veterinarians’ Training, Experience, and Comfort Level with Cases.”* (KOGAN et al., 2020)

- sich vordergründig mit human-psychologischen Fragen befassen und das Hauptaugenmerk nicht auf Hunde im Tierheim selbst gelegt ist.

Bsp.: *“Dog Training Intervention Shows Social-Cognitive Change in the Journals of Incarcerated Youth.”* (SYZMANSKI et al., 2018)

- sich mit Fragen der Mensch-Tier-Beziehung befassen und somit keine möglichst objektive, nur auf den individuellen Hund bezogene Beurteilung darstellt.

Bsp.: *“Relative efficacy of human social interaction and food as reinforcers for domestic dogs and hand-reared wolves.”* (FEUERBACHER et WYNNE, 2012)

- sich mit demographischen Aspekten bei der Abgabe von Hunden im Tierheim und/oder den Abgabegründen selbst befassen und daher nicht relevant für die tatsächliche spätere Beurteilung der Hunde sind.

Bsp.: *“A systematic review and meta-analysis of the proportion of dogs surrendered for dog-related and owner-related reasons.”* (LAMBERT et al., 2015)

- sich mit organisatorischen Hintergründen in Tierheimen befassen ohne Bezug zur Verhaltensbeurteilung.

Bsp.: *“Dogs on the Move: Factors Impacting Animal Shelter and Rescue Organizations’ Decisions to Accept Dogs from Distant Locations.”* (SIMMONS et HOFFMAN, 2016)

- sich mit konkreten klinisch-medizinischen Fragen oder der Messung klinischer Parameter bei Hunden im Tierheim durch Blut-/ Urin- oder Kotuntersuchungen befassen und sich daher nicht mit der vordergründig verhaltensmedizinischen Fragestellung dieser Arbeit beschäftigen.

Bsp.: *“Molecular detection and phylogenetic analysis of canine tick-borne pathogens from Korea.”* (SEO et al., 2020)

Bsp.: *“Daily fluctuation of urine serotonin and cortisol in healthy shelter dogs and influence of intraspecific social exposure.”* (ALBERGHINA et al., 2019)

- sich mit Katzen im Tierheim befassen, da Katzen kein Gegenstand dieser Diplomarbeit sind.

Bsp.: *“Perception of Cats: Assessing the Differences Between Videos and Still Pictures on Adoptability and Associated Characteristics.”* (SCHOENFELD-TACHER et al., 2019).

Auf dieselbe Art und Weise wurde mit allen in Tab. 2 genannten Suchbegriffskombinationen vorgegangen. Aufgrund der Tatsache, dass es sich bei PubMed® und weiteren Suchmaschinen um englischsprachige Meta-Datenbanken handelt, hatte die Suche mit deutschen Schlüsselwörtern geringen Erfolg – fast alle Suchergebnisse sind daher in Englisch. Nach Anwendung der Suchkriterien in den in Tab. 2 genannten Datenbanken wurden 90 Beiträge

gefunden. Sieben weitere Beiträge wurden in anderen Quellen gefunden. Es wurden somit initial 97 Studien zum Thema der Wesensbeurteilung im Tierheim nach Ausschluss unpassender Studien gefunden.

### 2.2.2 Vorauswahl

Nach Entfernung von Duplikaten ( $n = 1$ ) wurden insgesamt 96 Beiträge in die Vorauswahl aufgenommen. Diese 96 Beiträge wurden genauer betrachtet und mithilfe ihres Abstracts entweder ausgeschlossen oder zur weiteren Beurteilung des Volltextes eingeschlossen. Zur Beurteilung des Volltextes wurden Artikel aus wissenschaftlichen Fachzeitschriften (mit *peer review*-Verfahren) und Dissertationen einbezogen. Buchkapitel, Vorträge oder Übersichtsposter wurden aufgrund ihrer Fragwürdigkeit als wissenschaftliche Quelle nicht in die weitere Prüfung auf Eignung inkludiert – somit entfielen folgende Fundstücke:

- Bei zwei Beiträgen handelte es sich um Vorträge, zum einen für Freiwillige in einem Tierheim (SCHMORROW, 2020) und zum anderen für eine Tierschutzschulung (BENNETT, 2010, <https://www.youtube.com/watch?v=1QTEOjRRDCK>).
- Bei einem Beitrag handelte es sich um ein Kapitel aus einem Buch (WEISS et MOHAN, 2017).
- Ein Beitrag stellte sich als Poster/Übersicht heraus (GRAS, 2016).

In der Zusammenfassung lässt sich der Inhalt der Studien genauer ermitteln und somit wurden weitere, folgende Studien aufgrund von Verfehlen des Themas ausgeschlossen:

- In einer Studie stand die genetische Identifikation der Rassenzugehörigkeit von Tierheimhunden im Vordergrund, um die Übereinstimmung mit den durch das Tierheimpersonal „diagnostizierten“ Bezeichnungen zu eruieren (MARDER et VOITH, 2010). Hierbei wurden DNA-Analysen durchgeführt und keine Verhaltensbewertungen. Ähnliche Inhalte zur genetischen Codierung von Hunden im Tierheim wies die Arbeit von GUNTER et al., 2018 auf und wurde daher nicht berücksichtigt.
- In einer Studie (THIELKE et UDELL, 2019) wurden Hunde auf Pflegestellen und im Tierheim bezüglich ihrer Bindungsart und damit ihrer Sicherheit/Unsicherheit verglichen. Dies geschieht durch einen sogenannten *Secure Base Test*, und durch den Vergleich von C-BARQ- und MCPQ-R-Fragebögen. Die Vorteile von Pflegestellen sind durchaus ein relevantes Thema und weist Potential zu weiterer



Forschung auf, sind aber nicht Inhalt dieser Arbeit. Weiters ist die Evaluierung der Vermittelbarkeit Inhalt dieser Arbeit und weniger die der Bindungsstrukturen, da dies eher ein Kriterium von Lebensqualität darstellt.

Einige Arbeiten waren aufschlussreiche Literaturübersichten oder Reviews zum Thema, aus denen zwar durchaus Inhalte in diese Arbeit einfließen, allerdings wurden gezielt Studien zu diversen Beurteilungsverfahren auf die Kriterien diagnostischer Testsysteme untersucht, um diese dann vergleichen und auswerten zu können. Daher entfielen folgende Publikationen:

- DIEDERICH et GIFFROY, 2006: *“Behavioural testing in dogs: A review of methodology in search for standardisation“*.
- CLAY et al., 2020(c): *“In defense of canine behavioral assessments in shelters: Outlining their positive applications“*.
- JONES et GOSLING, 2005: *“Temperament and personality in dogs (Canis familiaris): A review and evaluation of past research“*.
- RAYMENT et al., 2015: *“Applied personality assessment in domestic dogs: Limitations and caveats“*.
- HAVERBEKE et al., 2015: *“Behavioral evaluations of shelter dogs: Literature review, perspectives, and follow-up within the European member states’s legislation with emphasis on the Belgian situation“*.
- PATRONEK et BRADLEY, 2016: *“No Better Than Flipping a Coin: Reconsidering Canine Behavior Evaluations in Animal Shelters“*.
- WIENER et HASKELL, 2016: *“Use of questionnaire-based data to assess dog personality“*.
- PATRONEK et al., 2019: *“What is the evidence for reliability and validity of behavior evaluations for shelter dogs? A prequel to “No better than flipping a coin““*.
- TAYLOR et MILLS, 2006: *“The development and assessment of temperament tests for adult companion dogs“*.

Somit wurden 16 Beiträge in der Vorauswahl ausgeschlossen.

### 2.2.3 Eignung

Achtzig Studien wurden anhand ihres Volltextes auf Eignung beurteilt.

Leider gab es einige Studien, von denen zwar ein Abstract, aber in weiterer Folge kein weiterer publizierter Volltext auffindbar war und welche darum zur weiteren Analyse nicht mehr in Frage kamen:

- *“Behaviour assessment tool may help shelter adoptions“* (KEIN AUTOR, 2007)
- *“Evaluation of sheltered dangerous dogs“* (MICHELAZZI et al. 2010)
- *“Shelter dog behavioral assessments: Behavioral coding and personality“* (MARDER et DOWLING-GUYER, 2010)
- *“Evaluation of a new screening tool for assessing temperament and behavior in dogs relinquished to shelters“* (KRUGER et al., 2009).

Für die Eignung kommen nur vollständige Manuskripte in Betracht. Die drei folgenden Beiträge waren Abstracts aus den Proceedings von Konferenzen. Die Abstracts wurden zwar in wissenschaftlichen Magazinen publiziert, es gab allerdings (noch) keine vollständigen Manuskripte zum Zeitpunkt der Veröffentlichung im Journal:

- *“Assessing shelter dogs for adoption suitability – an Australian perspective.“* (MORNEMENT et al., 2009) wurde später als *“A Review of Behavioral Assessment Protocols Used by Australian Animal Shelters to Determine the Adoption Suitability of Dogs.“* (MORNEMENT et al., 2010) publiziert.
- *“Validation of a new Temperament test as a practical tool for adoptions of sheltered dogs.“* (VALSECCHI et al., 2009) wurde später als *“Temperament test for re-homed dogs validated through direct behavioural observation in shelter and home environment.“* (VALSECCHI et al., 2011) publiziert.
- *“Human-directed aggression in shelter dogs: how to test for better prediction of outcomes“* (KLAUSZ et al., 2009) wurde später als *“A Quick Assessment Tool for Human-Directed Aggression in Pet Dogs“* (KLAUSZ et al., 2013) publiziert, allerdings zunächst ohne Tierheimbezug. KIS et al. veröffentlichten 2014 eine Folgestudie zur Anwendung im Tierheim.

Bei weiteren Studien waren Titel und Abstract zwar in englischer Sprache, der Volltext wurde allerdings weder in deutscher, noch in englischer Sprache verfasst und ist somit den Kriterien nicht entsprechend:

- Zwei Studien waren in spanischer Sprache: (BARRERA et al., 2013) und (JOSA et SUAREZ, 2008).

Es gab Studien, in denen zwar teilweise durch Ethogramme das Verhalten von Hunden in einer Tierheimsituation evaluiert wurde, allerdings nicht zum Zwecke der Wesensbeurteilung, sondern zum Zwecke der ethologischen Analyse der Reaktion auf die Tierheimumgebung, in Kombination mit diversen klinischen Parametern:

- HIBY et al. (2006) haben Verhaltensmuster in Verbindung mit dem Kortisolspiegel im Urin erfasst.
- PROTOPOPOVA (2016) hat sich mit Verhaltensmustern in Tierheimen befasst, mit besonderem Augenmerk auf Stereotypen, bedingt durch die Zwingerhaltung.
- GUNTER (2018) ging in einem Kapitel ihrer Dissertation auf Verhaltensweisen als Indikatoren für das Tierwohl im Tierheim ein.
- STEPHEN et LEDGER (2005) gingen in ihrer Arbeit ebenfalls auf Verhaltensmuster resultierend aus der Haltung im Tierheim ein und den damit verbundenen Einschränkung der Lebensqualität der Hunde.
- CAFAZZO et al. (2014) befassten sich mit Verhalten als Indikator für Tierwohl, gemeinsam mit einer Blutuntersuchung auf reaktive Sauerstoffmetaboliten.
- OWCZARCZAK-GARSTECKA et BURMAN (2016) gingen vorrangig auf Schlaf- und Ruheverhalten als Indikator für Tierwohl ein.
- BAUER et al. (2017) und MUGENDA et al. (2019) untersuchten den *Field Instantaneous Dogs Observation Test (FIDO)*, welcher auf die Beurteilung der Lebensqualität des Hundes abzielt.
- WALKER et al. (2016); ARENA et al. (2017); MENCHETTI et al. (2019) und ARENA et al. (2019) untersuchten *Qualitative Behavioural Assessment* und *Free Choice Profiling*, welches ebenfalls Verfahren sind, um die Lebensqualität von Hunden zu untersuchen und zu bewerten.

Es ist nicht ganz eindeutig, inwiefern die rein ethologische Fragestellung der Reaktion des Hundes auf eine Umgebung im Vordergrund steht oder die Wesensevaluierung selbst zum Zwecke der besseren Einschätzung des Hundes für die Vermittlung. So gab es in einigen Studien zwar Beobachtungen von Hunden in Situationen (und deren Aufzeichnungen mithilfe eines Ethogramms), welche denen in Wesenstests sehr ähnlich waren, die allerdings das Verhalten des Hundes als Reaktion auf seine Umgebung erfassen sollen und somit zur Einschätzung des Tierwohlfaktors dienen:

- *“Chronic stress in dogs subjected to social and spatial restriction I”* (BEERDA et al., 1999).
- *“Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs”* (TITULAER et al., 2013).
- *“Development and validation of a quality of life assessment tool for use in kennelled dogs (Canis familiaris)”* (KIDDIE et COLLINS, 2014).

Eine der Studien befasste sich mit verschiedenen Methoden zur Ermittlung von Prävalenzdaten zu Problemverhalten im Tierheim (NORMANDO et al., 2019). Die Studie wurde exkludiert, da das Ziel auch hier in der Beurteilung und Verbesserung der Lebensqualität von Hunden im Tierheim war und nicht die Beurteilung der Vermittelbarkeit.

Weitere Studien waren nicht geeignet, da Gruppen von Hunden beobachtet wurden, nicht aber der Einzelhund als Individuum:

- DALLA VILLA et al. (2013) beobachteten Gruppen von vier bis fünf Hunden mithilfe eines Ethogramms und sammelten Speichelproben zur Kortisolmessung.
- PETAK (2013) hat die sozialen Interaktionen von Hunden in einer Gruppenhaltung durch ein Ethogramm erfasst.

Eine der Studien ist zwar bereits älter als 25 Jahre, wurde aber dennoch mit einbezogen, da sie die erste Studie zum Thema gewesen sein dürfte (VAN DER BORG et al., 1991). Eine weitere Studie (CLAY et al., 2020b) wurde inkludiert, obwohl die Studienobjekte Hunde in Privatbesitz waren. Da aber der Fokus dennoch auf der Anwendung im Tierheim lag, wurde sie trotz dessen einbezogen.

In diesem Schritt wurden weiters jene Studien exkludiert, die zwar von großer Bedeutung sind für die Beurteilung der jeweiligen Testsysteme selbst, deren teilnehmende Hundepopulation allerdings in Privatbesitz ist und somit wiederum nicht die Tierheimsituation widerspiegelt:

- SERPELL et HSU (2001): *“Development and validation of a novel method for evaluating behavior and temperament in guide dogs.”*
- HSU et SERPELL (2003): *“Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs.”*
- NAGASAWA et al. (2011); TAMIMI et al. (2015); GONZÁLEZ-RAMÍREZ et al. (2017); CANEJO-TEIXERA et al. (2018) wendeten den C-BARQ in diversen Ländern an.
- BARNARD et al. (2012): *“Validity of model devices used to assess canine temperament in behavioral tests.”*
- DUFFY et SERPELL (2012): *“Predictive validity of a method for evaluating temperament in young guide and service dogs.”*
- LOYER (2019): *“Social Behavior and Cognitive Bias in the Domestic Dog.”*
- JONES (2008): *“Development and validation of a dog personality questionnaire.”*
- NETTO et PLANTA (1997): *“Behavioural testing for aggression in the domestic dog.”*
- PLANTA et De MEESTER (2007): *“Validity of the Socially Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans.”*
- De MEESTER et al. (2008) *“A preliminary study on the use of the Socially Acceptable Behavior test as a test for shyness/confidence in the temperament of dogs.”* und (2011) *“The use of the socially acceptable behavior test in the study of temperament of dogs.”*
- BENNETT et al. (2012): *“Investigating behavior assessment instruments to predict aggression in dogs.”*
- IVERSEN (2008): *„Ein Aggressionstest bei Berliner Stadthunden - Feststellung vorkommender Aggressionsformen, Reizschwellen und Einflussfaktoren für aggressives Verhalten.“*

- THIESEN-MOUSSA et al. (2018): „*Der Niedersächsische Wesenstest: Ergebnisse des Testens der Gefährlichkeit von Hunden.*“
- MIRKÒ et al. (2013): „*Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (Canis familiaris).*“

Das Mindestalter der Hunde, die in die jeweiligen Untersuchungen inkludiert wurden, schwankte zwischen vier Monaten (beispielsweise POULSEN et al., 2010), sechs Monaten (beispielsweise CLAY et al., 2019) und einem Jahr (beispielsweise VALSECCHI et al., 2011). Aufgrund dessen wurde ein Mindestalter der zu evaluierenden Hunde von vier Monaten festgesetzt, damit die Ergebnisse einheitlicher und damit aussagekräftiger sind:

- HENNESSY et al. (2001) betrachteten in ihrer Studie 166 Hunde, von denen 106 Hunde noch Welpen (Vorhandensein von Milchzähnen) waren, es wurde kein Mindestalter festgelegt. Leider war eine genauere Aufschlüsselung des Verhältnisses von Welpen zu adulten Hunden in den Ergebnissen nicht möglich. Die Studie wurde daher ausgeschlossen.

Somit wurden insgesamt 46 Beiträge in der Eignung ausgeschlossen. Aufgrund des Themenumfangs wurden die verbliebenen 34 Studien (siehe Tab. 3) den folgenden Unterpunkten zugeordnet und einzeln evaluiert:

- 1) Verhaltensevaluierung via Fragebögen
- 2) Verhaltensevaluierung via Wesenstest
- 3) Verhaltensevaluierung via Ethogramm

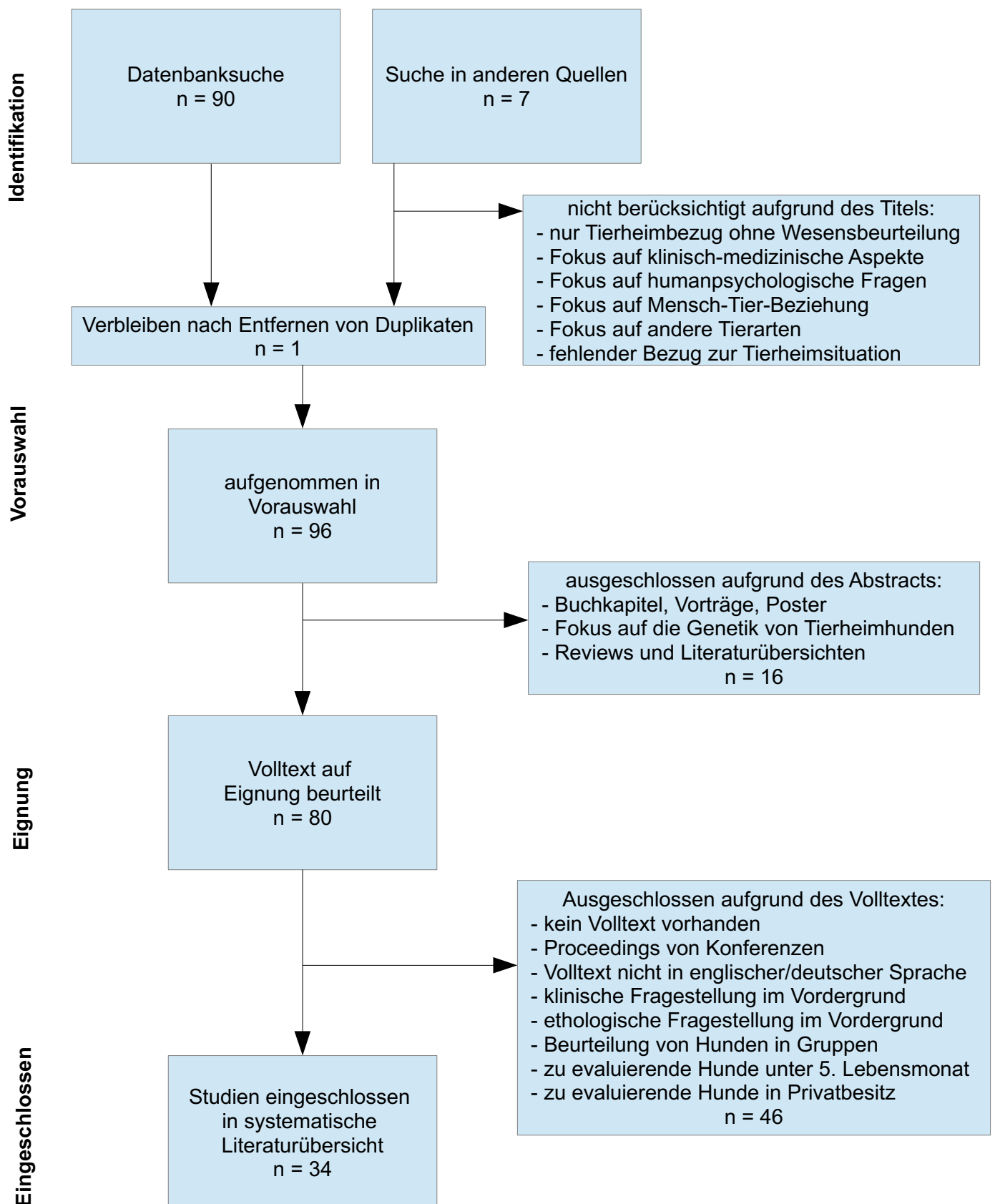


Abb. 1: Flussdiagramm zur Darstellung des Vorgehens bei der Literaturrecherche

Tab. 3: Übersicht der inkludierten Studien

Nr.	Studie	AutorInnen	Jahr der Publikation	Publikationsland	Name des Testverfahrens	Methode
1	Evaluation of a behavioural assessment questionnaire for use in the characterization of behavioural problems of dogs relinquished to animal shelters	Segurson et al.	2005	USA	C-BARQ	Fragebogen
2	Evaluation of a behavioural assessment tool for dogs relinquished to shelters	Duffy et al.	2014	USA	C-BARQ	Fragebogen
3	Relinquishing dog owners' ability to predict behavioural problems in shelter dogs post adoption	Stephen et Ledger	2007	United Kingdom		Fragebogen
4	Human and canine personality assessment instruments to predict successful adoptions with shelter dogs	Walker	2014	USA	MCPQ-R	Fragebogen
5	Behavioural testing of dogs in animal shelters to predict problem behaviour	van der Borg et a.	1991	Niederlande		Wesenstest
6	Aggressive behavior in adopted dogs that passed a temperament test	Christensen et al.	2007	USA	Assess-A-Pet	Wesenstest
7	Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs	Bollen et Horowitz	2008	USA	Assess-A-Pet	Wesenstest
8	Characteristics and Adoption Success of Shelter Dogs Assessed as Resource Guardians	McGuire	2019	USA	Assess-A-Pet	Wesenstest
9	Abilities of Canine Shelter Behavioral Evaluations and Owner Surrender Profiles to Predict Resource Guarding in Adoptive Homes	McGuire	2020	USA	Assess-A-Pet	Wesenstest
10	Service dog selection tests: Effectiveness for dogs from animal shelters	Weiss et Greenberg	1997	USA	SAFER™	Wesenstest
11	Selecting shelter dogs for service dog training	Weiss	2002	USA	SAFER™	Wesenstest
12	Preliminary Investigation of Food Guarding Behavior in Shelter Dogs in the United States	Mohan-Gibbons et al.	2012	USA	SAFER™	Wesenstest
13	Comparison of SAFER™ behaviour assessment results in shelter dogs at intake and after a 3-day	Bennett et al.	2015	USA	SAFER™	Wesenstest



	acclimation period					
14	The impact of excluding food guarding from a standardized behavioural canine assessment in animal shelters	Mohan-Gibbons et al.	2018	USA	SAFER™	Wesenstest
15	An evaluation of a behavioural assessment to determine the suitability of shelter dogs for rehoming	Poulsen et al.	2010	Australien	RSPCA	Wesenstest
16	Do Behavior Assessments in a Shelter Predict the Behavior of Dogs Post-Adoption?	Clay et al.	2020(a)	Australien	RSPCA	Wesenstest
17	Comparison of Canine Behaviour Scored Using a Shelter Behaviour Assessment and an Owner Completed Questionnaire, C-BARQ	Clay et al.	2020(b)	Australien	RSPCA, C-BARQ	Fragebogen, Wesenstest
18	Temperament test for re-homed dogs validated through direct behavioural observation in shelter and home environment	Valsecchi et al.	2011	Italien	Temperament Test	Wesenstest
19	Revisiting a Previously Validated Temperament Test in Shelter Dogs, Including an Examination of the Use of Fake Model Dogs to Assess Conspecific Sociability	Barnard et al.	2019	Italien	Temperament Test	Wesenstest
20	Development of the behavioral assessment for re-homing K9's (BARK) protocol	Mornement et al.	2014	Australien	B.A.R.K.	Wesenstest
21	Evaluation of the predictive validity of the Behavioral Assessment for Re-homing K9's (BARK) protocol and owner satisfaction with adopted dogs	Mornement et al.	2015	Australien	B.A.R.K.	Wesenstest
22	Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK	Diesel et al.	2008	United Kingdom	Dogs Trust	Wesenstest
23	Intra-rater and Inter-rater Reliability of the ASPCA's Behaviour Evaluation of Fearful Dogs	Matters	2016	USA	ASPCA®	Wesenstest
24	Behavioral traits detected in shelter dogs by a behavior evaluation	Dowling-Guyer et al	2011	USA	Match-Up	Wesenstest
25	Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption	Marder et al.	2013	USA	Match-Up	Wesenstest
26	Consistency of shelter dogs' behavior toward a fake versus real stimulus dog during a behaviour	Shabelansky et al.	2015	USA	Match-Up	Wesenstest

	evaluation					
27	Timing and presence of an attachment person affect sensitivity of aggression tests in shelter dogs	Kis et al.	2014	Ungarn		Wesenstest
28	Entwicklung und Anwendung eines Verhaltenstests für Tierheimhunde sowie die Untersuchung der Verhaltensentwicklung im Tierheim	Frank	2009	Deutschland		Wesenstest
29	Assessing adoptability in shelter dogs through a behavioural test	Zilocchi et al.	2015	Italien		Wesenstest
30	Ethotest: A new model to identify (shelter) dogs' skills as service animals or adoptable pets	Lucidi et al.	2005	Italien	Ethotest	Wesenstest
31	Qualified „in shelter“ dogs´ evaluation and training to promote successful dog-human relationships	Sticco et al.	2011	Italien	Ethotest	Wesenstest
32	Evaluating the temperament in shelter dogs	De Palma et al.	2005	Italien		Ethogramm
33	Modelling pesonality, plasticity and predictability in shelter dogs	Goold et Newberry	2017	United Kingdom		Ethogramm
34	Early Recognition of Behaviour Problems in Shelter Dogs by Monitoring them in their Kennels after Admission to a Shelter	Clay et al.	2019	Australien		Ethogramm

#### **2.2.4 Einteilung in Evidenzklassen**

Da es sich um ein verhaltensmedizinisches Themengebiet handelt, konnten keine Schemata der evidenzbasierten Medizin direkt angewandt werden. Die evidenzbasierte Medizin beschreibt den gewissenhaften, ausdrücklichen und rationalen Gebrauch der gegenwärtig besten wissenschaftlichen Evidenz für Entscheidungen, wie der individuelle Patient medizinisch versorgt wird (PSCHYREMBEL, 2020). Eines der Ziele ist die Identifizierung sicherer, präziser und wirksamer Untersuchungsverfahren.

Die Qualität von Studien lässt sich durch die Einteilung in Evidenzklassen beurteilen. Da die Einteilung in Evidenzklassen von verhaltensmedizinischen Studien allerdings nicht gängig ist, lehnte ich mich an die Einteilung von SANTESSO et al. (2006) an und wandelte sie ab (siehe Tab. 4). Folgende Aspekte wurden beurteilt bzw. für die Anwendbarkeit im Bereich Verhaltensstudien abgewandelt:

Die Erhebung einer Follow-Up-Studie ist Voraussetzung, um weiterführend statistische Validitätskriterien wie Sensitivität, Spezifität und prädiktive Werte zu erfassen.

Ein großes Problem der Wesenstests ist, dass Referenzwerte im neuen Zuhause zumeist durch die Euthanasie von Hunden nach Nicht-Bestehen des Wesenstests nur unvollständig erhoben werden können.

Die Befragung abgebender BesitzerInnen durch Fragebögen ist schwierig zu randomisieren, da einerseits nur ein Teil der Tierheimhundepopulation aus Abgabehunden besteht und andererseits nicht jede/r abgebende/r BesitzerIn in eine Befragung einwilligt. Daher spiegelt diese Methode der Verhaltensbeurteilung nur einen Teil der Tierheimhundepopulation wider und kann daher nicht als randomisierte Auswahl bezeichnet werden.

Unter einer randomisierten/ blinden Auswahl an Hunden verstehe ich im Zusammenhang mit Wesenstests entweder eine zufällige Auswahl von Hunden, welche einem Wesenstest unterzogen werden (bei Interventionsstudien) oder eine Erfassung aller Hunde, welche im Studienzeitraum einem Wesenstest unterzogen wurden (bei Beobachtungsstudien). Sobald Hunde gezielt ausgewählt wurden für eine Studie oder gewisse Hunde durch eine Vorselektion exkludiert wurden, kann nicht mehr von einer blinden Auswahl gesprochen werden.

Beobachtungsstudien sind jene Studien, in denen die jeweilige Verhaltensbeurteilung in den teilnehmenden Tierheimen ohnehin stattfindet und von den AutorInnen für einen gewissen Zeitraum untersucht und ausgewertet wird. Interventionsstudien sind jene Studien, in denen die jeweilige Verhaltensbeurteilung durch die AutorInnen initiiert wird zum Zweck der Studie.

Tab. 4: Einteilung der Studien in Evidenzklassen

Evidenzklasse	Anforderungen an die Studie
Platin	<ul style="list-style-type: none"> <li>• Mindestens zwei Kontrollgruppen</li> <li>• Stichprobengröße mindestens 50 Hunde pro Gruppe</li> <li>• Teilnahme an Follow-Up-Studie mindestens 80%</li> <li>• blinde Auswahl von Hunden</li> </ul>
Gold	<ul style="list-style-type: none"> <li>• Eine Kontrollgruppe, welche die o.g. Anforderungen erfüllt</li> </ul>
Silber	<ul style="list-style-type: none"> <li>• die o.g. Anforderungen werden nicht erfüllt</li> <li>• Follow-Up-Studie ist vorhanden, keine Datenverfälschung durch Euthanasie von Hunden, welche den Wesenstest nicht bestanden haben</li> <li>• blinde Auswahl von Hunden</li> </ul>
Bronze	<ul style="list-style-type: none"> <li>• Follow-Up-Studie ist durch Euthanasie der Hunde, welche den Wesenstest nicht bestanden haben, verfälscht ODER</li> <li>• keine Follow-Up-Studie vorhanden</li> <li>• Auswahl an Hunden wurde vorselektiert, nicht zufällig</li> </ul>

## 3 Ergebnisse

### 3.1 Einteilung der Studien in Evidenzklassen

Lediglich zwei der Studien hatten mehr als eine Kontrollgruppe, jedoch fehlte in diesen beiden Studiendesigns die Randomisierung (SEGURSON et al., 2005; DUFFY et al., 2014) und es wurden keine Angaben darüber gemacht, ob Hunde aufgrund der Testergebnisse euthanasiert wurden (siehe Tabelle 6). In einer der beiden Studien wurde kein Follow-Up durchgeführt (SEGURSON et al., 2005) und in der anderen Studie lag die Teilnahme an ebendieser bei nur 12% (DUFFY et al., 2014). Daher konnte keine der Studien der Platinklasse zugeordnet werden.

Auch der Goldklasse konnte keine der Studien zugeordnet werden. Alle Studien, außer den beiden oben genannten, haben zwar eine Kontrollgruppe, aber keine dieser Studien erfüllte alle Anforderungen der Platinklasse. In neun der Studien war die Stichprobengröße kleiner als die Mindeststichprobengröße von 50 Hunden (WEISS et GREENBERG, 1997; BENNETT et al., 2015; ZILOCCHI et al., 2015; MATTERS, 2016; DIESEL et al., 2008; KIS et al., 2014; LUCIDI et al., 2005; STICCO et al., 2011; CLAY et al., 2019). In acht der verbleibenden Studien wurden die Hunde nicht randomisiert ausgewählt (STEPHEN et LEDGER, 2007; WALKER, 2014; VAN DER BORG et al., 1991; WEISS, 2002; MOHAN-GIBBONS, 2012; CLAY et al., 2020b; BARNARD et al., 2019; MORNEMENT et al., 2014) und in zwei weiteren Studien fehlen Angaben zur Auswahl der Hunde (FRANK, 2007; VALSECCHI et al., 2011). In drei der verbleibenden Studien gab es keine Follow-Up-Studie (DOWLING-GUYER et al., 2011; DE PALMA et al., 2005; GOOLD et NEWBERRY, 2017) und in acht Studien lag die Teilnahme an ebendieser unter dem Mindestwert von 80% (CLAY et al., 2020a; MOHAN-GIBBONS et al., 2018; POULSEN et al., 2010; CHRISTENSEN et al., 2007; BOLLEN et HOROWITZ, 2008; MCGUIRE et al., 2020; MARDER et al., 2013; MORNEMENT et al., 2015).

Vier Studien können der Silberklasse zugeordnet werden, da sie zwar nicht die Anforderungen an Gold und Platin erfüllen, jedoch über eine randomisierte Auswahl an Hunden und eine Follow-Up-Studie verfügen, welche nicht durch Euthanasie von positiv-getesteten Hunden verfälscht wurde (MCGUIRE, 2019; MCGUIRE et al., 2020; CLAY et al., 2019; LUCIDI et al., 2005). Sämtliche andere Studien sind der niedrigsten Evidenzklasse (Bronze) zuzuordnen (siehe Tabelle 6).

In Tab. 5 findet sich ergänzend eine Übersicht der vorgenommenen Follow-Up-Studien inklusive deren Zeitpunkte.

Tab. 5: Anzahl und Zeitpunkt der Follow-Up Studien – Übersicht

Studie	Anzahl Follow-Up´s	Zeitpunkt(e) Follow-Up post-Adoption
Mohan-Gibbons et al., 2012	3	nach 3 Tagen, 3 Wochen, 3 Monaten
Sticco et al., 2011	ca. 12	nach 2 Wochen, dann alle 2 Monate für ein Jahr
Stephen et Ledger, 2007	2	nach 2 und 6 Wochen
Clay et al., 2020a	1	nach 1 Monat
Van der Borg et al., 1991	1	nach 1 bis 2 Monaten
McGuire et al., 2020	1	nach 1 bis 3 Monaten
Duffy et al., 2014; Walker, 2014	1	nach 2 Monaten
Mornement et al., 2014 und 2015	1	nach 2 bis 8 Monaten
Poulsen et al., 2010	1	nach ca. 80 Tagen
Marder et al., 2013	1	nach mindestens 3 Monaten
Valsecchi et al., 2011	1	nach 4 Monaten
Bollen et Horowitz, 2008	1	nach 6 Monaten
Christensen, 2007	1	innerhalb von 13 Monaten
Lucidi et al., 2005	1	nach 14 Monaten PatientInnenkontakt

Tab. 6: Einteilung der inkludierten Studien in Evidenzklassen

Studie	Kontrollgruppen	Stichprobengröße	Randomisierung	Studientyp	Euthanasie von positiven Hunden	Follow-Up-Studie	Teilnehmeranzahl an der Follow-Up-Studie	Evidenzklasse
Segurson et al., 2005	2	54	-	Interventionsstudie	k. A.	-	kein Follow-Up	Bronze
Duffy et al., 2014	3	438	-	Interventionsstudie	k. A.	+	12%	Bronze
Stephen et Ledger, 2007	1	163	-	Interventionsstudie	+	+	59%	Bronze
Walker, 2014	1	197	-	Interventionsstudie	-	+	100%	Bronze
Van der Borg et al., 1991	1	81	-	Interventionsstudie	-	+	89%	Bronze
Christensen et al., 2007	1	279	+	Beobachtungsstudie	+	+	24%	Bronze
Bollen et Horowitz, 2008	1	2017	+	Beobachtungsstudie	+	+	27%	Bronze
McGuire, 2019	1	1015	+	Beobachtungsstudie	-	(+) Rückgabe ja/nein	100%	Silber
McGuire et al., 2020	1	139	+	Beobachtungsstudie	-	+	37%	Silber
Weiss et Greenberg, 1997	1	9	-	Interventionsstudie	-	(+) Reevaluierung nach Ausbildung	100%	Bronze
Weiss, 2002	1	75	-	Interventionsstudie	-	(+) Reevaluierung nach Ausbildung	53%	Bronze
Mohan-Gibbons et al., 2012	1	96	-	Beobachtungsstudie	-	+	63%	Bronze
Bennett et al., 2015	1	33	+	Interventionsstudie	+	-	kein Follow-Up	Bronze
Mohan-Gibbons et al., 2018	1	14180	+	Kombination	+	(+) Rückgabe ja/nein	44-54%	Bronze
Frank, 2007	1	70	k.A.	Interventionsstudie	-	-	kein Follow-Up	Bronze
Poulsen et al., 2010	1	236	+	Beobachtungsstudie	+	+	19%	Bronze

Clay et al., 2020 (a)	1	955	+	Beobachtungsstudie	+	+	13%	Bronze
Clay et al., 2020 (b)	1	107	-	Interventionsstudie	-	(+) BesitzerInnenbefragung vorab	100%	Bronze
Valsecchi et al., 2011	1	78	k.A.	Interventionsstudie	-	+	44%	Bronze
Barnard et al., 2019	1	50	-	Interventionsstudie	-	-	kein Follow-Up	Bronze
Mornement et al., 2014	1	102	-	Interventionsstudie	+	+	71%	Bronze
Mornement et al., 2015	1	102	+	Interventionsstudie	+	+	72%	Bronze
Zilocchi et al., 2015	1	26	k.A.	Interventionsstudie	-	(+) Vermittlung ja/nein	100%	Bronze
Matters, 2016	1	12	-	Beobachtungsstudie	+	-	kein Follow-Up	Bronze
Dowling-Guyer et al., 2011	1	668	+	Beobachtungsstudie	+	-	kein Follow-Up	Bronze
Marder et al., 2013	1	376	+	Beobachtungsstudie	+	+	26%	Bronze
Diesel et al., 2008	1	20	k.A.	Beobachtungsstudie	-	-	kein Follow-Up	Bronze
Kis et al., 2014	1	25	+	Interventionsstudie	-	-	kein Follow-Up	Bronze
Lucidi et al., 2005	1	23	+	Interventionsstudie	-	+ Reevaluierung nach PatientInnenkontakt	30%	Silber
Sticco et al., 2011	1	32	k.A.	Interventionsstudie	-	+	38%	Bronze
De Palma et al., 2005	1	74	+	Interventionsstudie	-	-	kein Follow-Up	Bronze
Goold et Newberry, 2017	1	3263	+	Interventionsstudie	-	-	kein Follow-Up	Bronze
Clay et al., 2019	1	38	+	Interventionsstudie	-	(+) Wesenstest als Referenzwert	100%	Silber



## **3.2 Verhaltensbeurteilung durch Fragebögen**

Da eine BesitzerInnenbefragung das Vorhandensein von BesitzerInnen voraussetzt, ist die Anwendbarkeit im Tierheimbereich auf eine Befragung von VorbesitzerInnen bei der Abgabe von Hunden beschränkt. Fundhunden und beschlagnahmten Hunden bleibt diese Möglichkeit daher naturgemäß verwehrt.

Die Wesensbeurteilung über einen Fragebogen wurde erstmals durch SERPELL und HSU im Jahr 2001 publiziert.

### **3.2.1 Fragebögen in der Praxis**

#### **3.2.1.1 Verhaltensbeurteilung durch C-BARQ und Modifikationen**

Der Fragenkatalog "C-BARQ" (*Canine Behavioral Assessment & Research Questionnaire*) ist ein standardisiertes Instrument für TierpflegerInnen, HundebesitzerInnen, HundeführerInnen und für HundexpertInnen, um Verhalten zu evaluieren (UNIVERSITY OF PENNSYLVANIA, 2020). Den Beginn von C-BARQ stellte eine Studie zur Auswahl von Blindenführhunden und Diensthunden im Alter von einem Jahr dar (SERPELL und HSU, 2001). Im Jahre 2003 wurde der C-BARQ von HSU und SERPELL weiterentwickelt, um die Prävalenz und den Schweregrad von Verhaltensproblemen bei Hunden in Privatbesitz und bei Arbeitshunden zu messen. Ziel dieser Methode sollte es sein, Hunde mit Verhaltensproblemen von Hunden mit normalem Verhalten zu differenzieren und Hunde für ihren möglichen Arbeitseinsatz vorzuselektieren.

In mehreren Ländern wurde der C-BARQ bereits angewendet, um das Verhalten von Hunden zu beurteilen, so wurden beispielsweise im Iran (TAMIMI et al., 2015) und in Mexiko (GONZÁLEZ-RAMÍREZ et al., 2017) BesitzerInnen befragt. Auch in Japan (NAGASAWA et al., 2011) und Portugal (CANEJO-TEIXERA et al., 2018) wurde der C-BARQ aufgegriffen und auf seine Fähigkeit untersucht, Verhaltensprobleme bei Hunden in Privatbesitz zu identifizieren.

SEGURSON et al. haben im Jahre 2005 erstmals versucht, einen Fragebogen bei Eingang der Hunde im Tierheim zu evaluieren. In einer Querschnittsstudie wurden 54 Abgabehunde und 784 Hunde bei TierbesitzerInnen beurteilt und verglichen. Die sieben Messparameter wurden in insgesamt 45 Fragen im C-BARQ abgefragt.

Es handelt sich um eine prospektive Studie, in der die Hypothese lautete: Abgebende BesitzerInnen könnten geneigt sein, bei der Angabe ernsthafter Problemverhalten nicht ganz ehrlich zu sein, was den Fragebogen weniger sinnvoll, wenn nicht gar potenziell gefährlich macht. Daher wurden die abgebenden BesitzerInnen in zwei Gruppen geteilt:

1. BesitzerInnen, denen die Vertraulichkeit der Daten zugesichert wurde
2. BesitzerInnen, deren Daten offiziell auch für die Weitervermittlung verwendet wurden.

Diese beiden Datensätze wurden im Folgenden miteinander verglichen und der vertrauliche Datensatz wurde weiters mit den Hunden in Privatbesitz verglichen, um die Prävalenz bestimmter Problemverhalten in der Hundepopulation messen zu können. Diese Studie misst daher die Objektivität von *C-BARQ*.

Auch bei DUFFY et al. (2014) handelt es sich um eine prospektive Studie zur Anwendung des *C-BARQ*. In 42 Fragen wurden insgesamt 23 Messparameter abgefragt. Die Befragten hatten eines von drei eindeutigen Vorworten zu lesen und wurden somit in drei Gruppen eingeteilt:

1. BesitzerInnen, denen erklärt wurde, dass die Antworten vom Tierheimpersonal zur weiteren Evaluierung der Hunde für die Vermittlung verwendet werden (nicht-vertrauliche Gruppe)
2. BesitzerInnen, denen versichert wurde, dass die Antwort nicht vom Tierheimpersonal gesehen werden (vertrauliche Gruppe)
3. BesitzerInnen, denen nichts zur Verwertung der Antworten gesagt wurde (Kontrollgruppe).

Untersucht wurden außerdem die interne Konsistenz und die Fähigkeit des *C-BARQ*, zwischen Hunden entsprechend ihres späteren Schicksals zu unterscheiden (Adoption oder Euthanasie). Weiters wurde untersucht, wie sehr die Beurteilung der VorbesitzerInnen mit der Beurteilung des Tierheimpersonals bezüglich der Aggressivität eines Hundes übereinstimmt. Außerdem wurde die Korrelation der *C-BARQ*-Ergebnisse mit dem Verhalten der Hunde etwa zwei Monate nach Adoption verglichen. Die NeubesitzerInnen hatten die Standardversion des *C-BARQ* (mit 100 Fragen) auszufüllen.

MCGUIRE et al. (2020) haben in ihrer Befragung der VorbesitzerInnen Teile des *C-BARQ*-Fragebogens modifiziert und angewendet. In dieser Version wurden nur drei Messparameter

in acht Fragen erfragt. Die Befragung der abgebenden BesitzerInnen erfolgte hierbei zunächst zu allgemeinen Gründen der Abgabe und im Folgenden speziell zu Ressourcenverteidigung und aggressivem Verhalten. Wurde mindestens einmal mit „Knurren“, „Schnappen“ oder „Beißen“ geantwortet, wurden die Befragung als positiv für Ressourcenverteidigung gewertet. Es folgte eine Befragung, ob der Hund jemals jemanden angeknurrt, geschnappt oder gebissen hat, jeweils mit Bitte um Beschreibung der Situation. Sechszwanzig Prozent der Hunde mit VorbesitzerInnen wurden aufgrund von Verhaltensproblemen abgegeben. Knapp 30% berichteten einen Vorfall bezüglich Ressourcenverteidigung in der Vergangenheit. Über 53% berichteten, dass der Hund bereits mindestens einmal Zeichen von Aggression gezeigt hat. Die AutorInnen haben alle drei Informationsquellen (VorbesitzerInnen, Wesenstest und NeubesitzerInnen) zusammengefügt und verglichen.

In der Studie von CLAY et al. (2020b) wurden die Ergebnisse der BesitzerInnenbefragung durch C-BARQ verglichen mit den Ergebnissen des Wesenstests der RSPCA während eines Besuches im Tierheim. Eine Abgabe der Hunde im Tierheim stand dabei nicht im Raum, es handelte sich um Hunde in Privatbesitz.

Somit wurden für den C-BARQ Untersuchungen zu Objektivität (SEGURSON et al., 2004; DUFFY et al., 2014), zur internen Konsistenz (DUFFY et al., 2014) und zur prognostischen Validität (DUFFY et al., 2014; MCGUIRE et al., 2020) vorgenommen (siehe Tab. 7).

Tab. 7: Messung der diagnostischen Testkriterien im C-BARQ-Fragebogen, angewandt an Tierheimhunden

Fragebogen	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
C-BARQ	✓	-	✓	-	-	-	-	✓

### 3.2.1.2 Fragebogen nach STEPHEN et LEDGER 2007

STEPHEN et LEDGER (2007) haben ebenfalls einen Fragebogen entwickelt und in einer Studie publiziert. In dieser Studie wurden Berichte der VorbesitzerInnen über das Verhalten ihrer Hunde mit dem Verhalten dieser Hunde im neuen Zuhause nach zwei und sechs Wochen verglichen. Der Fragebogen besteht aus 20 Fragen und erfragt zehn typische Problemverhalten. Einhundertdreiundsechzig VorbesitzerInnen haben bei der Abgabe ihres Hundes im Tierheim den Fragebogen ausgefüllt, von diesen abgegebenen Hunden wurden 99 Hunde neu vermittelt. Die restlichen Hunde wurden entweder in andere Unterkünfte überstellt oder aus medizinischen/ verhaltenstechnischen Gründen euthanasiert. Die NeubesitzerInnen füllten den Fragebogen nach zwei Wochen und nach sechs Wochen aus. Der Vergleich der beiden Befragungen nach Adoption wurde als Test-Retest-Reliabilität untersucht. Neben der Test-Retest-Reliabilität wurde die Vorhersagekraft des Fragebogens der VorbesitzerInnen für das Verhalten der Hunde nach der Adoption untersucht, wofür die Befragung der NeubesitzerInnen als Referenzwert galt.

Tab. 8: Messung der diagnostischen Testkriterien im Fragebogen nach STEPHEN et LEDGER

Fragebogen	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
nach STEPHEN et LEDGER	-	-	-	✓	-	-	-	✓



### 3.2.2 Vergleichende Übersicht aller Fragebögen

In der folgenden Tabelle (Tab. 10) findet sich eine Übersicht über die gemessenen diagnostischen Kriterien, aufgeschlüsselt auf die einzelnen Studien. Leider misst keine der Studien alle relevanten Kriterien, ebenfalls wurden für keinen Fragebogen all diese Kriterien gemessen (siehe Tabellen 6 bis 8).

Tab. 10: Übersicht über die untersuchten diagnostischen Kriterien in den Studien mit Tierheimbezug

Studie	Anzahl Messparameter	Anzahl Einzelfragen	Objektivität			Reliabilität			Validität		
			standardisierter Fragebogen	Anzahl BeurteilerInnen	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität
Segurson et al., 2005 (C-BARQ)	7	45	✓	1	✓						
Duffy et al., 2014 (C-BARQ)	23	42	✓	1	✓		✓				✓
McGuire et al., 2020 (C-BARQ)	3	8	✓	1							✓
Clay et al., 2020b (C-BARQ)	10	102	✓	1							
Stephen et Ledger, 2007	10	20	✓	1				✓			✓
Walker, 2014 (MCPQ-R)	5	26	✓	1							✓

Der Fokus der Fragebögen liegt zum einen auf möglichen Alltagsproblemen (beispielsweise *Aggression gegenüber fremden Menschen* in STEPHEN et LEGDER, 2007; SEGURSON et al., 2005; DUFFY et al., 2014; CLAY et al., 2020b), aber auch auf Verhaltensweisen, die das zukünftige Arbeiten mit dem Hund beeinflussen könnten (beispielsweise *Aufmerksamkeit bei*

*Kommandos* in STEPHEN et LEDGER, 2007; *Trainingsfokus* in WALKER, 2014; *Trainierbarkeit* in DUFFY et al., 2014). In der folgenden Tabelle (Tabelle 11) findet sich eine Übersicht der gemessenen Messparameter, welche in den jeweiligen Studien berücksichtigt wurden.

Tab. 11: Messparameter, welche in den Fragebögen berücksichtigt wurden

Studie	Messparameter										
	Aggression ggü. Menschen	Angst/ Furcht <sup>2</sup>	Artgenossen	Trennungsangst	Aufmerksamkeit/ Bemühen um Aufmerksamkeit	Erregung/ Nervosität	Trainierbarkeit / Gehorsam	Ziehen an der Leine	Bellen	Freundlichkeit / Umgänglichkeit	weitere
Segurson et al., 2005 (C-BARQ)	Fremde/ BesitzerIn	Fremde/ nichtsozial	Angst/ Aggression	+	+						
Duffy et al., 2014 (C-BARQ)	Fremde/ BesitzerIn	Fremde/ nichtsozial	Angst/ Rivalität/ Aggression	+	+	+	+	+	+		+
McGuire et al., 2020 (C-BARQ)	+										+
Clay et al., 2020b (C-BARQ)	+	+		+	+	+	+				+
Stephen et Ledger, 2007	+	+			+	+			+		+
Walker, 2014 (MCPQ-R)					+		+			+	

<sup>2</sup> Einige AutorInnen haben Angst unterteilt in Angst im Allgemeinen (*anxiety*) und Furcht vor bestimmten Reizen (*fear*), so beispielsweise STEPHEN et LEDGER (2007).



### 3.2.3 Diagnostische Testkriterien der Fragebögen

#### 3.2.2.1 Objektivität von Fragebögen

Abgebende BesitzerInnen könnten voreingenommen sein (SEGURSON et al., 2005). Daher war die Hauptfrage, die sich SEGURSON et al. (2005; *C-BARQ*) stellten, die Frage nach der Objektivität der Beurteilung durch die VorbesitzerInnen. Im Vergleich der beiden Abgabegruppen miteinander (*vertraulich* vs. *nicht-vertraulich*) wichen die Messparameter „*Aggression gegen den/die BesitzerIn oder gegen Familienmitglieder*“ und „*Angst gegenüber Fremden*“ signifikant voneinander ab: die vertrauliche Gruppe erzielte in diesen Messparametern einen höheren („schlechteren“) Punktescore als die nicht-vertrauliche Gruppe. Die abweichenden Antworten der VorbesitzerInnen nach unterschiedlicher Vorinformation weisen auf eine Beeinflussbarkeit der BeurteilerInnen und somit auf einen groben Mangel hinsichtlich der Objektivität hin.

Die Information, die den abgebenden BesitzerInnen im Vorwort des Fragebogens gegeben wurde, hatte in der Studie von DUFFY et al. (2014; *C-BARQ*) allerdings keinen einheitlichen Einfluss auf die gegebenen Antworten. Dies widersprach somit den Ergebnissen von SEGURSON et al. (2005). Nur zwischen den drei teilnehmenden Tierheimen gab es in manchen Punkten signifikante Abweichungen, nicht allerdings hinsichtlich *Vertraulichkeit*.

#### 3.2.2.2 Reliabilität von Fragebögen

STEPHEN und LEDGER (2007; *Fragebogen nach STEPHEN et LEDGER*) untersuchten die Test-Retest-Reliabilität mithilfe von zwei Befragungen nach Adoption. Hierfür füllten 259 NeubesitzerInnen den Fragebogen nach zwei Wochen und 199 NeubesitzerInnen nach sechs Wochen aus. Für 191 Hunde gab es jeweils zwei Fragebögen. Die Übereinstimmung der beiden Befragungen der NeubesitzerInnen war moderat bis hoch (z.B. niedrigste Korrelation  $r = 0.46$  für *Aufmerksamkeit während Spaziergängen*; stärkste Korrelation  $r = 0.73$  für *Angst/ Aggression gegenüber der Tierärztin/ dem Tierarzt*). Alle Übereinstimmungen waren signifikant ( $P < 0.001$ ).

Im Kontext von Fragebögen beruft sich die Reliabilität vor allem auf die interne Konsistenz (WIENER et HASKELL, 2016), welche allerdings nur DUFFY et al. (2014; *C-BARQ*)

gemessen haben. Alle Kategorien hatten eine mindestens akzeptable interne Konsistenz (*Cronbachs Alpha*  $\geq 0.70$ ), außer „*nicht-soziale Angst*“ ( $\alpha = 0.625$ ), diese war fraglich.

### 3.2.2.3 Validität von Fragebögen

Untersucht wurde durch DUFFY et al. (2014) die Fähigkeit des *C-BARQ*, zwischen Hunden entsprechend ihres späteren Schicksals zu unterscheiden (Adoption oder Euthanasie). Hunde, welche zur Adoption freigegeben wurden, zeigten signifikant bessere Punktescores bezüglich „*Aggression gegenüber Fremden*“. Demnach unterschieden die Ergebnisse der VorbesitzerInnenbefragung laut der Studie erfolgreich die Hunde bezüglich ihres schlussendlichen Ausgangs (Adoption oder Euthanasie), was hinweisend für eine konvergente Validität ist. Es wird von der BesitzerInnenbefragung erwartet, dass sie mit dem Ausgang für den Hund korreliert: von schlechteren Punktescores wird eine Korrelation zur Tendenz zur Euthanasie erwartet.

Als Referenzwert für die Beurteilung der prognostischen Validität von Fragebögen kann beispielsweise das Ergebnis eines Wesenstests dienen oder besser noch das Verhalten nach der Adoption.

So wurde durch DUFFY et al. (2014; *C-BARQ*) untersucht, wie sehr die Beurteilung der VorbesitzerInnen mit der Beurteilung des Tierheimpersonals bezüglich der Aggressivität eines Hundes übereinstimmt und somit das Ergebnis des Eingangstests voraussagt. In zwei der drei teilnehmenden Tierheimen nahm das Personal routinemäßig eine Wesensbeurteilung bei Eingang der Hunde vor. Hierfür wurden folgende binäre Variablen eingeführt und darauffolgend mit den jeweiligen *C-BARQ*-Ergebnissen verglichen:

- *Hunde, die beim Eingangstest in irgendeiner Form Aggression gezeigt haben*
- *Hunde, die beim Eingangstest in keinerlei Form Aggression gezeigt haben*

Hunde, die beim Eingangstest in irgendeiner Form Aggression gezeigt haben, erhielten im *C-BARQ* in allen vier Aggressionskategorien, sowie in „*Angst gegenüber Fremden*“, „*Angst gegenüber Hunden*“ und „*Schwanz jagen*“ signifikant schlechtere Punktescores. Die Antworten der VorbesitzerInnen bezüglich *Aggression* und *Angst gegenüber Menschen und anderen Hunden* korrelierten mit der unabhängigen subjektiven Beurteilung der Aggressivität durch das Tierheimpersonal in einem Wesenstest.

Weiters wurde die Korrelation der C-BARQ-Ergebnisse durch DUFFY et al. (2014) mit dem Verhalten der Hunde etwa zwei Monate nach Adoption verglichen. Die NeubesitzerInnen hatten die Standardversion des C-BARQ (mit 100 Fragen) auszufüllen. Auch hier wurden signifikante Korrelationen in mehreren Kategorien („*Aggression gegenüber Fremden*“  $r_s = 0.494$ ,  $P < 0.001$ ; „*Kauen ungeeigneter Gegenstände*“  $r_s = 0.402$ ,  $P < 0.01$ , „*Urinieren bei Alleinlassen*“  $r_s = 0.421$ ,  $P < 0.01$ ) zwischen den Aussagen der Vor- und NeubesitzerInnen festgestellt.

Die Beschreibungen der abgegebenen BesitzerInnen zu Ressourcenverteidigung in MCGUIRE et al. (2020) sagten das Verhalten im neuen Zuhause durch den C-BARQ signifikant voraus. Der positive prädiktive Wert zum Vorhandensein von Ressourcenverteidigung durch die Informationen der abgebenden BesitzerInnen lag bei 38.5%, der negative prädiktive Wert bei 93.6%. Das Vorhandensein von Verhaltensproblemen oder Aggression laut Aussage der VorbesitzerInnen war kein signifikanter Voraussagewert für das Vorkommen von Ressourcenverteidigung im neuen Zuhause. MCGUIRE et al. haben alle drei Informationsquellen (VorbesitzerInnen, Wesenstest und NeubesitzerInnen) zusammengefügt, allerdings gab es nur bei 44 Hunden alle Informationen. Bei 65.9% der Hunde stimmten alle drei Informationsquellen überein, was in der Studie als ausreichend bezeichnet wurde. Diese Ergebnisse sprechen gegen ein „*under-reporting*“ der VorbesitzerInnen bezüglich Ressourcenverteidigung.

Durch CLAY et al. (2020b) wurden 107 Hunde einmal durch ihre BesitzerInnen im C-BARQ und ein weiteres Mal durch die AutorInnen in einem Wesenstest beurteilt, ein bis zwei Wochen später. Korrelationen wurden in den Messparametern *Angst* (z.B.  $r = 0.34$  für *Angst* im Wesenstest und *Angst gegenüber Fremden* im C-BARQ,  $P < 0.01$ ), *Freundlichkeit* (z.B.  $r = -0.32$  für *Freundlichkeit* im Wesenstest und *Angst gegenüber Fremden* im C-BARQ,  $P < 0.01$ ) und *Aggression* (z.B.  $r = 0.32$  für *Aggression* im Wesenstest und *Berührungssensitivität* im C-BARQ,  $P < 0.01$ ) gefunden. Die getestete Population spiegelt allerdings keine reine Tierheimhundepopulation wider, sondern bestand aus Hunden in Privatbesitz. Nur 44.9% der Hunde waren ursprünglich aus dem Tierheim. Den Referenzwert für die prognostische Validität stellte ein Wesenstest bei einem Tierheimbesuch dar.

Zusammenfassend zum C-BARQ kann man sagen, dass die Vorhersagekraft für das Verhalten im neuen Zuhause maximal „mäßig“ ist (DUFFY et al., 2014). Während der negative prädiktive Wert für *Ressourcenverteidigung* hoch war, lag der positive prädiktive Wert bei nur 38.5%, was für sehr viele falsch-positive Ergebnisse spricht (MCGUIRE et al., 2020).

Für 96 Hunde in der Studie von STEPHEN et LEDGER (2007; *Fragebogen nach STEPHEN et LEDGER*) gab es jeweils einen Fragebogen von Vor- und NeubesitzerIn. Es wurde die Vorhersagekraft des *Fragebogens nach STEPHEN et LEDGER* für das Verhalten der Hunde nach der Adoption untersucht, wofür die Befragung der NeubesitzerInnen als Referenzwert galt. Die Übereinstimmung zwischen der Befragung der VorbesitzerInnen und die der NeubesitzerInnen war nur in neun von 20 Punkten signifikant (*Angst vor der Tierärztin/ dem Tierarzt, Furcht in der Tierarztpraxis oder beim Alleinbleiben, Kauen an Möbeln, sexuelles Besteigen, Aggression gegen fremde Hunde/ fremde Menschen/ gegenüber der Tierärztin/ dem Tierarzt, Stehlen von Essen*), siehe Tab. 12. Die Übereinstimmung der Erstbefragung mit der Befragung nach zwei Wochen post Adoption war höher (8/20 Fragepunkte) als mit der Befragung nach sechs Wochen post Adoption (6/20 Fragepunkte). Bei fünf von 20 Fragepunkten hat sich das Verhalten zwischen Vor- und NeubesitzerIn signifikant verändert (*Kauen an Möbeln, Stehlen von Essen, exzessives Bellen, Aufmerksamkeit bei den Kommandos „Sitz“ und „Bleib“*). Die Befragung der VorbesitzerInnen hat sich in dieser Studie für den Großteil der Fragepunkte als nicht effektiv zur Vorhersage von Verhalten im neuen Zuhause erwiesen, so erreichten beispielsweise die Messparameter *Angst gegenüber Fremden/ fremden Hunden/ lauten Geräuschen, Aufregung und Aufmerksamkeit* keine signifikanten Übereinstimmungen.

Alle fünf Messparameter bei WALKER (2014) im *MCPQ-R-Fragebogen* ergaben eine schwache Korrelation zwischen den Beurteilungen von Vor- und NeubesitzerInnen. Die Korrelation reichte von  $r = -0.02$  (*Motivation*) bis  $r = 0.21$  (*Neurotizismus*). Für *Extroversion* lag die Korrelation bei  $r = 0.15$ , für *Trainingsfokus* bei  $r = 0.08$  und für *Freundlichkeit* bei  $r = 0.15$ . Die Übereinstimmungen waren dennoch signifikant, das heißt nicht zufällig ( $P = 0.02$ ). Es besteht ein Zusammenhang, aber dieser ist schwach. *Trainingsfokus* und *Freundlichkeit* wurden durch die VorbesitzerInnen als weniger stark ausgeprägt beurteilt als durch die NeubesitzerInnen.

In der folgenden Tabelle (Tab. 12) lassen sich die statistischen Validitätskriterien vergleichen. Im direkten Vergleich sollte allerdings berücksichtigt werden, dass die AutorInnen die statistischen Messgrößen nicht einheitlich erfasst haben. So haben beispielsweise STEPHEN et LEDGER (2007; *Fragebogen nach STEPHEN et LEDGER*) und DUFFY et al. (2014; *C-BARQ*) den *Spearman'schen Rangkorrelationskoeffizienten* angewandt, um die Übereinstimmung zu ermitteln, während WALKER (2014; *MCPQ-R*) den *Konkordanzkorrelationskoeffizienten* angewandt hat. MCGUIRE et al. (2020; *C-BARQ*) haben dagegen die prädiktiven Vorhersagewerte gemessen.

Tab. 12: Statistische Validitätskriterien (die Beurteilung der prognostischen Validität ist nur dann möglich, wenn eine Follow-Up-Studie stattfand)

Studie	Messparameter							
	Aggression gegenüber Menschen	Angst/ Furcht	Artgenossen	Trennungsangst	Trainierbarkeit/ Gehorsam	Erregung/ Aufregung	Freundlichkeit	Ressourcen-verteidigung/ Futteraggression
Duffy et al., 2014 (C-BARQ)	<i>Fremde:</i> mäßige Korrelation $r_s = 0.494$ (P<0.001)	keine signifikanten Korrelationen	keine signifikanten Korrelationen	keine signifikanten Korrelationen	keine signifikanten Korrelationen	keine signifikanten Korrelationen		
McGuire et al., 2020 (C-BARQ)								PPW 38.5% NPW 93.6%
Stephen et Ledger, 2007	<i>Fremde:</i> moderate bis deutliche Korrelation $r_s = 0.43$ (P<0.001) <sup>3</sup> $r_s = 0.73$ (P<0.001)  <i>Tierarzt/ Tierärztin:</i> moderate Korrelation $r_s = 0.44$ (P<0.01) $r_s = 0.48$ (P<0.1)	<i>Tierarzt/ Tierärztin:</i> moderate Korrelation $r_s = 0.44$ (P<0.01) $r_s = 0.46$ (P<0.1)	<i>Aggression ggü. fremden Hunden:</i> moderate Korrelation $r_s = 0.43$ (P<0.001) $r_s = 0.53$ (P<0.01)	deutliche Korrelation $r_s = 0.71$ (P<0.001) $r_s = 0.65$ (P<0.001)	keine signifikanten Korrelationen	keine signifikanten Korrelationen		
Walker, 2014 (MCPQ-R)					schwache Korrelation $\rho_c = 0.08$ (P=0.02)		schwache Korrelation $\rho_c = 0.15$ (P=0.02)	

3 Wert 1 gibt die Korrelation zwei Wochen nach Adoption an, Wert 2 gibt die Korrelation sechs Wochen nach Adoption an

### 3.3 Verhaltensbeurteilung durch Wesenstests

Das Leben im Tierheim ist für Hunde oft stresserfüllt und traumatisch aufgrund von sensorischer Überstimulation, sozialer Isolation, der Änderung der Routine, Kontrollverlust und einer völlig neuen Umgebung. Stress hat weitreichende Folgen und kann die kognitiven Fähigkeiten, das Verhalten und den emotionalen Zustand des Hundes beeinflussen. Die meisten Hunde haben zum Zeitpunkt der Aufnahme in ein Tierheim keinerlei Erfahrung mit der neuen Situation und verfügen daher kaum über positive Bewältigungsstrategien. Manche Hunde verfallen in negative Bewältigungsstrategien, beispielsweise Meideverhalten, Hemmung oder Beschwichtigung, aber auch Aggression (CLAY et al., 2020a).

Daher sollte der Ort *Tierheim* und dadurch bedingte Frustration, Erregung, Reizüberflutung, Deprivation und Stress in eine Beurteilung einbezogen werden. In dieser Umgebung können die Reaktionen von Hunden nicht mit Reaktionen in stabiler und geregelter Umgebung gleichgesetzt werden. Hinzu kommt, dass auch der Zeitpunkt der Verhaltensevaluierung eine große Rolle spielen könnte, besonders in Zusammenhang mit dem Stresslevel (BENNETT et al, 2015).

Die Erwartungen an Verhaltensbeurteilungen im Tierheim (sogenannte „Wesenstests“) sind somit sehr hoch, folgende Ziele werden dabei verfolgt (STERNBERG, 2017):

- besseres Zusammenpassen von Hunden und NeubesitzerInnen
- Abklärung der Vermittelbarkeit – neutralere Bewertung als aufgrund der Rasse, der Aufenthaltsdauer etc.
- Aggressionspotenzial detektieren inkl. der Reizschwellen für aggressives Verhalten
- soll sichereren Umgang mit Tierheimhunden gewährleisten
- soll bessere Möglichkeit der Beschreibung gewährleisten
- soll maßgeschneidertes Training an Stärken und Schwächen des einzelnen Hundes gewährleisten
- soll bessere Einschätzung der notwendigen Unterbringung gewährleisten
- soll bessere Aufklärung der NeubesitzerInnen über Problemvermeidung, Training,... gewährleisten

- soll Hunde herausdetektieren, die nicht sicher zur Vermittlung/zur Betreuung durch Freiwillige sind
- soll das Risiko für die öffentliche Sicherheit minimieren durch die Identifikation von Hunden mit Neigung zu Aggression

Grundsätzlich werden Wesenstests durch Einzeltests aufgebaut und meist durch eine von zwei Methoden beurteilt (DOWLING-GUYER et al., 2011). Die erste Methode ist die Verhaltenscodierung (*behavioral coding*), die sich auf individuelle Verhaltensweisen fokussiert und deren Anwesenheit, Frequenz, Dauer und Latenzzeit bestimmt. Ein Beispiel hierfür ist, wie ein Hund auf Fremde zugeht. Springt er hoch oder beschnuppert er die Person? „*Anspringen ja/nein*“ wäre eine Möglichkeit der dichotomen Datenerfassung. Wie oft er hochspringt an der Person, wäre eine Möglichkeit der numerischen Datenerfassung. Diese Methode wurde von VAN DER BORG et al. (1991); WEISS et GREENBERG (1997); WEISS (2002); CLAY et al. (2020b); DOWLING-GUYER et al. (2011); MARDER et al. (2013) und DIESEL et al. (2008) angewandt. Die zweite Methode sind Verhaltensbewertungsskalen (*behavioral rating scales*). Hierfür wird Verhalten nach definierten Messparametern gruppiert (beispielsweise *Aggression*) und anschließend wird jedes Level genau definiert, angelehnt an einer Eskalationsleiter. Ein Beispiel hierfür ist eine 5-Punkte-Skala, bei der „1“ für *keine Aggression*, „2“ für *Knurren oder Bellen*, „3“ für *Zähne zeigen*, „4“ für *Schnappen* und „5“ für *Beißen oder Attackieren mit Beißintention* steht. Alle übrigen in dieser Literaturübersicht inkludierten Studien, welche den Großteil der Studien darstellen, haben die Bewertung des Wesenstests mit dieser Methode durchgeführt.

Typische Einzeltests in einem Wesenstest sind beispielsweise: Spielen mit einem Objekt, Reaktion auf einen unbekanntem Raum, Reaktion auf einen Artgenossen, ein Puppentest zur Evaluierung der Reaktion auf Kinder, Streicheln des Hundes, das Erteilen von Grundkommandos, ein bedrohliches Annähern, die Reaktion auf das Öffnen eines Regenschirmes in unmittelbarer Nähe des Hundes, der Test auf das Vorhandensein von Ressourcenverteidigung oder das Verhalten an der Leine (TAYLOR et MILLS, 2006).

Es wurde bereits eine Vielzahl von Wesenstests entwickelt, modifiziert und auf ihre Anwendbarkeit untersucht. Im Folgenden werden jene Wesenstests vorgestellt, zu denen es im Zuge der Literaturrecherche Publikationen zur Anwendung im Tierheim gab.



### 3.3.1 Wesenstests in der Praxis

#### 3.3.1.1 Wesenstests in Europa

##### 3.3.1.1.1 Wesenstest *nach VAN DER BORG et al., 1991*

VAN DER BORG et al. (1991) haben in den Niederlanden einen Wesenstest entwickelt, um die Rückgaberate von Hunden ins Tierheim zu vermindern. Der Test sollte einfach genug sein, um vom Tierheimpersonal selbst durchgeführt werden zu können und beinhaltet verschiedene mögliche Alltagssituationen und Stimuli, die Problemverhalten verursachen können. In der Studie sollte der Vorhersagewert gemessen werden, indem die Ergebnisse des Wesenstests mit Interviews der NeubesitzerInnen, sowie der Einschätzung des Tierheimpersonals verglichen wurden. Getestet wurden nur jene Hunde, die geeignete Kandidaten zur Vermittlung waren. Auch sollte die Varianz der Abgabegründe möglichst hoch sein. Daher war die Auswahl nicht zufällig.

Im Jahre 1997 wurde der Test durch NETTO et PLANTA mit Unterstützung durch das Ministerium für Landwirtschaft, Umweltschutz und Fischerei weiterentwickelt. Der neue Test bestand aus über 80 Teiltests und das primäre Ziel sollte der Zuchtausschluss von besonders aggressiven Individuen sein. In der Pilotstudie wurden keine Tierheimhunde, sondern Hunde mit BesitzerInnen auf dem Testgelände der Universität Utrecht vorgestellt, teilweise mit Beißvorgeschichte. Anfangs beginnt der Wesenstest mit unbedrohlichen Teiltests, wie einigen Unterordnungsübungen im Freien und später werden die Teiltests provokativer, beispielsweise mit einer fremden Testperson, welche den/die BesitzerIn beschimpft und anschreit. Die Auswertung erfolgt nach einer *5-Punkte-Gutman-Skala* (5 = *Beißen und/oder Attackieren mit Beißintention*), wobei die Detektion von Hunden mit Tendenzen zum höchsten Punkt der Skala das wichtigste Ziel des Tests darstellen soll. Der *Dutch Kennel Club* hat bei den AutorInnen einen weiteren, gekürzten Test in Auftrag gegeben (PLANTA et De MEESTER, 2007). Dieser sogenannte *Socially Acceptable Behavior test* (*SAB Test*) umfasst 16 Einzeltests zur Evaluierung der Aggressivität von Hunden. Leider konnten keine Studien zur Anwendung dieser beiden Tests im Tierheim gefunden werden.

Somit wurde für den Wesenstest nach VAN DER BORG et al. lediglich die Vorhersagekraft untersucht, siehe Tab. 13.

Tab. 13: Messung der diagnostischen Testkriterien im Wesenstest nach VAN DER BORG et al.

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
VAN DER BORG	(✓)	-	-	-	-	-	-	✓

### 3.3.1.1.2 Ethotest© und Modifikationen

Zwar basieren die Evaluierungsmethoden in diesem Beurteilungssystem auf Verhaltensbeobachtungen, die Abläufe entsprechen aber eher einem Wesenstest, weshalb sie an dieser Stelle eingeordnet werden.

LUCIDI et al. haben 2005 den Ethotest© entwickelt, mit dem Ziel Hunde aus dem Tierheim als geeignete Therapiehunde zu identifizieren. Im Jahre 2011 wurde dieser Ethotest© durch STICCO et al. noch einmal aufgegriffen und vereinfacht, um die generelle Eignung von Tierheimhunden für hypothetische AdoptantInnen zu testen.

Im Ethotest© wird ein Algorithmus in drei Schritten angewandt. In Schritt A werden die Aggressivität und das Temperament erfasst, wofür ein Binärsystem verwendet wird (*ja/nein*). Die „gutmütigen“ und unterwürfigen Hunde bestehen Schritt A. Schritt B misst die Fähigkeiten in der Interaktion mit Menschen. Die Reaktionen des Hundes wurden für eine festgelegte Dauer von einer Minute beobachtet und notiert. Dieser Schritt muss positiv absolviert werden. Schritt C testet die Fähigkeit des Hundes, auf einfache Kommandos von (mindestens drei) verschiedenen Personen zu reagieren. Dies soll Aussage über seine Trainierbarkeit treffen. Der Ethotest© soll:

- aggressive Hunde aus der tiergestützten Therapie fernhalten

- Hunde mit der richtigen Eignung identifizieren und die Auswahl auf jene beschränken, die durchwegs auf Kommandos reagieren
- ermöglichen, dass alle Hunde jeden Geschlechts und jeder Rasse/Mischung in einem Alter über ein Jahr einbezogen werden
- jene Hunde identifizieren, die für die Arbeit mit verschiedenen PartnerInnen geeignet sind.

In dieser Studie wurden 23 Mischlingshunde mit einer Aufenthaltsdauer von sieben Tagen bis zwei Jahren im Tierheim getestet. Acht von 23 Hunden haben Schritt A nicht bestanden. Sechs der 15 verbleibenden Hunde haben Schritt B nicht bestanden. Drei der restlichen neun Hunde haben Schritt C nicht bestanden. Somit haben insgesamt sechs Hunde den Ethotest© bestanden. Zusätzlich zu den sechs Hunden, die durch den Ethotest© identifiziert wurden, hat die *Delta Society*<sup>4</sup> noch zwei weitere Hunde für geeignet befunden. Das Ergebnis der *Delta Society* galt als Referenzwert. Einer der sechs Hunde, die den Ethotest© bestanden haben, konnte durch die *Delta Society* nicht evaluiert werden, da er bereits vermittelt war zu dem Zeitpunkt. Somit gingen in Summe sieben Hunde in die Therapiearbeit.

Neben der Auswahl von Therapiehunden soll der Ethotest© auch die Vermittelbarkeit steigern, da die Vermittlungsrate von Hunden, die als „*geeignet für die Therapiearbeit*“ befunden wurden, erhöht sein soll aufgrund der Tatsache, dass dies für besondere Umgänglichkeit spricht.

STICCO et al. (2011) haben in ihrer Studie den Ethotest© vereinfacht (*Ethotest-R*), um mit ihm die Eignung von 32 Hunden zu testen, den Erwartungen von hypothetischen NeubesitzerInnen zu entsprechen. Die Hunde in der Studie befanden sich bereits zwischen zwei Monaten und sechs Jahren im Tierheim. Test A testet Aggressivität und Umgänglichkeit. In diesem Teil können Hunde direkt aus der Vermittlung ausscheiden, sofern sie Aggressivität zeigen. Neun von 32 Hunden haben diesen Teil nicht geschafft. Test B testet die Ängstlichkeit, Test C testet verschiedene Reaktionen auf Stimuli und Test D testet die Ansprechbarkeit auf Training und Spiel. Die Verhaltensweisen wurden mit Punkten bewertet, zum Bestehen ist ein gewisser Punktescore notwendig. Wird dieser nicht erreicht, gilt der Hund als (noch)

---

4 Unter dem Verband *Delta Society* (heute *Pet Partners*®) schlossen sich im Jahre 1977 Organisationen, WissenschaftlerInnen, TiertrainerInnen, TierpsychologInnen und TierbesitzerInnen zusammen und führten Standards und Richtlinien für den Einsatz von Tieren in der Therapie ein (<https://petpartners.org/>)

nicht geeignet für die Vermittlung. Vier der verbleibenden 23 Hunde haben die notwendige Gesamtpunktzahl nicht erreicht. Nur zwölf von 22 Hunden fanden ein neues Zuhause, einer von ihnen lag unter dem Mindestpunktescore.

Somit wurde auch für den Ethotest© nur die prognostische Validität untersucht.

Tab. 14: Messung der diagnostischen Testkriterien im Ethotest© und Ethotest-R

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
Ethotest© und Ethotest-R	(✓)	-	-	-	-	-	-	✓

### 3.3.1.1.3 Wesenstest nach FRANK, 2007

Die Autorin FRANK entwickelte einen Wesenstest, der als Eingangstest (durchzuführen innerhalb von fünf Tagen) Verhaltensprobleme aufdecken soll und die Vermittelbarkeit beurteilen soll. Der Test soll nach mindestens vier Wochen wiederholt werden, um eine Aussage über die Verhaltensentwicklung treffen zu können. FRANK hat hierzu 70 Hunde im Eingangstest geprüft, 23 davon nach vier Wochen im Wiederholungstest. Der Wesenstest misst zwar insgesamt 14 Messparameter, allerdings scheinen die wesentlichen Messparameter *Aggression* und *Vermittelbarkeit* zu sein.

Die Beurteilung erfolgt in fünf Leveln nach NETTO et PLANTA (1997) entsprechend der gezeigten Zeichen von Aggression. Die Skala ist die 5-Punkte-Skala nach GUTMANN.

- *Level 1*: keine Aggression
- *Level 2*: Knurren und/oder Bellen
- *Level 3*: Zähne zeigen mit/ohne Knurren und/oder Bellen

- *Level 4*: Schnappen (Bissbewegung aus der Distanz) mit/ohne Knurren und/oder Bellen und/oder Zähne zeigen, mit/ohne Annäherung (allerdings Stoppen in einer gewissen Distanz vom Stimulus)
- *Level 5*: Beißen oder Attackieren mit Beißintention; schnelle, direkte Annäherung ohne Stoppen, mit/ohne Knurren und/oder Bellen und/oder Zähne zeigen

Entsprechend des höchsten Levels, das der Hund mindestens einmal im Wesenstest gezeigt hat, wird er in die entsprechende Kategorie 1 bis 5 eingeteilt, welche die Vermittelbarkeit repräsentieren soll und über das weitere Vorgehen entscheidet. Hunden, welche der Kategorie 1 zugeordnet werden, steht einer direkten Vermittlung an neue BesitzerInnen nichts im Weg. Wird bei Hunden im Wesenstest ein Verhalten des Levels 2 beobachtet, fallen diese in Kategorie 2 und gelten ebenfalls als vermittelbar. Dennoch werden an dieser Stelle bereits intensive Vorgespräche bei der Vergabe und, nach Möglichkeit, die Teilnahme an Trainingsprogrammen angeraten. Zeigt der Hund im Wesenstest mindestens einmal ein Verhalten nach Level 3, fällt er in Kategorie 3 und darf nur an erfahrene HalterInnen unter Auflage einer Teilnahme an einem Trainingsprogramm vermittelt werden. Hunde der Kategorie 4 und 5 gehen nicht in die Vermittlung, sondern in intensive Verhaltenstherapien und Trainingsprogramme. Auch hier entspricht das Level des gezeigten Verhaltens der Zuordnung in die jeweilige Kategorie.

Bei acht Hunden (11.4%) musste der erste Wesenstest abgebrochen werden (teilweise war das Betreten des Zwingers nicht möglich, teilweise war das Anleinen des Hundes oder das Verlassen des Zwingers nicht möglich). Diese Hunde wurden Kategorie 5 zugeordnet. Einundfünfzig Hunde (72.9%) zeigten keinerlei aggressives Verhalten im gesamten Test. Diese Hunde wurden Kategorie 1 zugeordnet. Elf Hunde (15.7%) zeigten Verhalten der Level 2 bis 4, der Test musste aber nicht abgebrochen werden. Davon wurden zwei Hunde (2.9%) der Kategorie 4 zugeordnet werden und somit ebenfalls aus der Vermittlung ausgeschlossen werden. Die vier Hunde (5.7%), die in Kategorie 2 zugeordnet wurden, wurden somit ebenfalls als „vermittelbar“ eingeordnet. Zählt man die fünf Hunde (7%) der Kategorie 3 dazu, wären insgesamt 60 Hunde (85.7%) vermittelbar und zehn Hunde (14.3%) unvermittelbar aufgrund der Einordnung in Kategorie 4 oder 5. FRANK geht detailliert auf die Reaktionen der einzelnen Hunde in den Einzeltests ein.

Die Ergebnisse von Test und Wiederholungstest sind zwar vergleichbar, aber es wurde nicht explizit eine Test-Retest-Reliabilität ermittelt. Die prognostische Validität lässt sich lediglich bezüglich des Vermittlungsstatus nach vier Wochen ermitteln, es gibt kein Follow-Up im neuen Zuhause.

Tab. 15: Messung der diagnostischen Testkriterien im Wesenstest nach FRANK

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
nach FRANK	(✓)	-	-	(✗)	-	-	-	✓

#### 3.3.1.1.4 Wesenstest nach *Dogs Trust*

In der Studie von DIESEL et al. (2008) wurden die Inter-Rater-Reliabilität und Intra-Rater-Reliabilität eines Wesenstests untersucht. Dieser wird von der britischen Tierschutzorganisation *Dogs Trust* verwendet, ist aber (noch) nicht standardisiert zwischen den verschiedenen Tierheimen. Der Fokus der Studie liegt auf 40 Mitgliedern der Tierschutzorganisation *Dogs Trust*, welche das Verhalten von 20 Hunden auf Videoaufnahmen beurteilen sollen. Die Videoaufnahmen wurden in einem der 17 *Dogs Trust* Tierheime aufgenommen. Der Wesenstest besteht aus drei verschiedenen Situationen. Für die Videoaufnahmen wurde der Testablauf für alle 20 Hunde weitestgehend standardisiert.

Die Test-Retest-Reliabilität wurde nicht gemessen, allerdings wird in der Studie erwähnt, dass eine hohe Intra-Rater-Reliabilität umso wichtiger ist, da die Hunde in den Wesenstests für gewöhnlich nur einmal getestet werden.

Tab. 16: Messung der diagnostischen Testkriterien im Wesenstest von *Dog's Trust*

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
Dog's Trust	-	-	-	-	✓	✓	-	-

### 3.3.1.1.5 Wesenstest nach KLAUSZ et al., 2013

Eine initiale Studie zum Wesenstest nach KLAUSZ et al. (KLAUSZ et al., 2013) bezog sich auf Hunde in Privatbesitz. Um die prognostische Validität für Aggression des Wesenstests zu untersuchen, wurden die Beißvorgeschichte und die Auskünfte der BesitzerInnen als Referenzwert verwendet und die Hunde in drei Gruppen geteilt: *Nicht-Beißer*, *Einmal-Beißer* und *Mehrfach-Beißer*. In dieser Studie bestand der Wesenstest aus fünf Einzeltests (*freundliche Begrüßung*, *Wegnehmen eines Knochens*, *bedrohliche Annäherung*, *Tauziehen* und *auf-den-Rücken-rollen*). Die drei Einzeltests *freundliche Begrüßung*, *Wegnehmen eines Knochens* und *bedrohliche Annäherung* wurden weiters als effektiv darin beschrieben, spezifisches Verhalten entsprechend der Beißvorgeschichte hervorzurufen. Für *Tauziehen* und *auf-den-Rücken-rollen* ergaben sich keine signifikanten Unterschiede zwischen den drei Testgruppen. Diese beiden Einzeltests wurden daher in der zukünftigen Studie im Wesenstest exkludiert.

KIS et al. (2014) haben in ihrer Studie zum o.g. Wesenstest den Einfluss des Zeitpunkts des Wesenstests und die Anwesenheit der menschlichen Bezugsperson im Wesenstest untersucht. Hierbei wurde bereits der Tierheimkontext berücksichtigt. Der wesentliche Messparameter ist *Aggression*. Ziel der Studie ist es, die niedrige Sensitivität von Wesenstests durch die Untersuchung dieser Varianten zu erhöhen, damit weniger Falsch-Negative in der Vermittlung landen und demnach im neuen Zuhause Aggression zeigen,

obwohl sie im Wesenstest unauffällig waren. Zwei weitere Messparameter stellen *Angst/Unterwürfigkeit* und *Furcht/Unwohlsein* dar. Der Wesenstest beinhaltet nur noch drei Einzeltests (*Annäherung, Futter wegnehmen, Bedrohung*).

KIS et al. (2014) haben die Test-Retest-Reliabilität und die Inter-Rater-Reliabilität gemessen.

Tab. 17: Messung der diagnostischen Testkriterien im Wesenstest *nach KLAUSZ et al.*

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
<i>nach KLAUSZ et al.</i>	(✓)	-	-	✓	✓	-	-	-

### 3.3.1.1.6 *Temperament-Test*

VALSECCHI et al. (2011) haben es sich zur Aufgabe gemacht, die Validität und Reliabilität von einem Wesenstest, dem sogenannten *Temperament-Test*, zu untersuchen. Die Konsistenz der Reaktionen in Zeit und Situation wurde ebenfalls untersucht. Auch die interne Konsistenz wurde untersucht. Bei aggressivem Verhalten oder fehlender Reaktion auf positive Stimuli wurden niedrige Punktescores vergeben und bei „erwünschtem“ Verhalten (*freundlich, selbstbewusst, spielerisch, gehorsam*) wurden hohe Punktescores vergeben. Die Hunde wurden jeweils zweimal im Tierheim und einmal in ihrem neuen Zuhause vier Monate nach der Vermittlung getestet. Die Studie wurde in vier Phasen abgehalten. Insgesamt wurden 163 Hunde getestet. Einige dieser Hunde wurden mindestens 20 Tage nach Aufnahme im Tierheim getestet (*short-term*) und einige Hunde wurden erst mindestens ein Jahr nach durchgängigem Tierheimaufenthalt getestet (*long-term*). Da die meisten Hunde Fundhunde waren, konnten kaum Angaben zur Vorgeschichte erfasst werden. Es wurden die Kriteriumsvalidität (*criterion validity*) und die Inhaltsvalidität (*content validity*) untersucht.



Die Praktikabilität und Reproduzierbarkeit des Wesenstests nach VALSECCHI et al. stand in der Folgestudie von BARNARD et al. (2019) im Vordergrund. Der italienische *Temperament Test* wurde in zwei Tierheimen im Vereinigten Königreich angewendet, um die Reproduzierbarkeit zu evaluieren. Beide Tierheime sind sogenannte „no-kill-shelters“. Getestet wurden insgesamt 50 Hunde. Der Test nach VALSECCHI et al. (2011) wurde kaum verändert, mit Ausnahme von zusätzlichen *Fakehunden* (lebensechte Hundepuppen) als Stimulus.

Erneut wurden Inhaltsvalidität und interne Konsistenz beurteilt und mit den Ergebnissen von VALSECCHI et al. verglichen. Hierbei wurden weiters demographische Faktoren einbezogen. Für den *Temperament Test* wurden somit beinahe alle diagnostischen Testkriterien untersucht, siehe Tab. 18.

Tab. 18: Messung der diagnostischen Testkriterien im *Temperament Test*

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
Temperament Test	(✓)	✓	✓	✓	✓	✓	✓	-

### 3.3.1.1.7 Wesenstest nach ZILOCCHI et al., 2015

In dieser Studie wurde die Vermittelbarkeit von 26 Hunden evaluiert. Ziel sollte es sein, Problemhunde umgehend einem Verhaltensmodifikationsprogramm zuzuführen, um ihre Vermittelbarkeit zu steigern. Aufgrund der „no-kill-Politik“ in Italien sollen dadurch Hunde davor bewahrt werden, ihr gesamtes restliches Leben im Tierheim zu verbringen.

In Schritt 1 wurden Eigenschaften des Hundes evaluiert, die nicht veränderlich sind durch ein Erziehungsprogramm. Dazu zählen Alter, Größe, Fellfarbe und Morphologie. Auch bekannte Attacken auf Menschen oder Hunde in der Vergangenheit wurden einbezogen. In Schritt 2

wurden Informationen zum Verhalten in der täglichen Zwingeroutine vom Tierheimpersonal erfragt. Dabei wurde das Verhalten bei Annäherung durch Personen beurteilt, sowie das Verhalten bei Betreten und Verlassen des Zwingers und beim Spaziergang. In Schritt 3 wurde die Reaktion des Hundes bei direkter Annäherung einer fremden Person an den Zaun des Zwingers erfasst. Die Art der Annäherung ist genau definiert. Zwölf mögliche Reaktionsweisen wurden aufgeführt. In Schritt 4 wurde ein Wesenstest durchgeführt. Jeder Einzeltest wurde nach der Reaktion des Hundes beurteilt (*positive/ neutrale/ ruhige Reaktion; ängstliche Reaktion; aggressive Reaktion*), diese drei möglichen Reaktionen wurden als Messparameter gewertet. Anhand der Ergebnisse der vier Schritte wurde ein „Vermittelbarkeitsindex“ (*dog adoptability index - DAI*) berechnet, welcher die individuelle Vermittelbarkeit repräsentieren sollte. Kategorie 1 repräsentiert sofort vermittelbare Hunde (DAI ab 80), Kategorie 2 Hunde mit kritischer Vermittelbarkeit, aber ohne Aggression (DAI 60-79) und Kategorie 3 schwer vermittelbare Hunde, welche der Verhaltenstherapie bedürfen (DAI<60). Die getesteten Hunde bewegten sich in einem DAI von 50.5 bis 92.5. Im Detail wurden 57.7% (15 Hunde) Kategorie 1 zugeordnet, 25.9% (7 Hunde) wurden Kategorie 2 zugeordnet und 16.4% (4 Hunde) wurden Kategorie 3 zugeordnet. Leider kann der Studie nicht entnommen werden, über welchen Zeitraum hinweg die Hunde vermittelt wurden und ob ihnen entsprechend ihrer Kategorie Training zukam.

Für diesen Wesenstest wurde die prognostische Validität gemessen, da angegeben wurde, welche Hunde entsprechend ihrer Zuordnung zu einer der drei Kategorien vermittelt wurden.

Tab. 19: Messung der diagnostischen Testkriterien im Wesenstest nach ZILOCCHI et al.

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
nach ZILOCCHI et al.	(✓)	-	-	-	-	-	-	✓

### **3.3.1.2 Wesenstests in den USA**

#### **3.3.1.2.1 Assess-A-Pet**

SUE STERNBERG hat diesen Wesenstest im Jahre 2006 veröffentlicht. Die Ziele sollen zunächst die Abklärung der Vermittelbarkeit und in weiterer Folge eine bessere Passung zwischen Hund und NeubesitzerIn sein. Das Hauptaugenmerk liegt auf der Eruiierung der Reizschwelle für Aggression. Sowohl die Hunde, als auch die InteressentInnen werden in drei Level eingeteilt, je nach Reizschwelle der Hunde und Kompetenz der InteressentInnen. Getestet werden die Präsentation des Hundes im Zwinger, dessen Freundlichkeit und ob er Aggression in irgendeiner Form zeigt. Der Messparameter in *Assess-A-Pet* ist daher das Vorhandensein von Aggression, eingeteilt in vier Klassen: Aggression im Umgang und aus Frustration, Ressourcenverteidigung, Aggression gegenüber Fremden und Aggression gegenüber Hunden (STERNBERG, 2017).

In der Studie von CHRISTENSEN et al. (2007) sollte zum einen die Zahl der Falsch-Negativen evaluiert werden und zum anderen stellten sich die AutorInnen die Frage, ob relevante Parameter wirklich gut genug simuliert werden können. Siebenundsechzig Hunde wurden hierfür im Tierheim getestet, vermittelt und innerhalb von 13 Monaten nach Adoption telefonisch reevaluiert. Jeder positiv getestete Hund (Zeigen von Attackieren, Knurren, Schnappen, und/oder Beißen, mit Ausnahme von Ressourcenverteidigung), wurde als unvermittelbar erklärt und euthanasiert. Hunde, welche lediglich potentiell aggressives Verhalten (Starren, steife Körperhaltung, Bellen) zeigten, bestanden den Test. In die Vermittlung kamen nur jene Hunde, die den Test bestanden hatten, keine (bekannten) Bissvorfälle in ihrer Geschichte vorweisen konnten und keines dieser Verhalten im Tierheim gezeigt haben.

BOLLEN et HOROWITZ (2008) haben mithilfe des *Assess-A-Pet*-Wesenstests 2017 Tierheimhunde untersucht. Es sollten aggressive Tendenzen identifiziert werden und Verhaltensprobleme nach der Adoption vorausgesagt werden. Untersucht wurde auch die Verbindung zwischen dem Scheitern am Wesenstest, demographischen Faktoren des Hundes und dessen Vorgeschichte bezüglich Verhalten. Abgebende BesitzerInnen hatten ein Profil auszufüllen über demographische Daten und Informationen zum Verhalten des Hundes, inkl. offener Fragen zum Beschreiben von Verhaltensweisen. Der Wesenstest

erfolgte zumeist nach ca. 48 Stunden im Tierheim. Der einzige relevante Messparameter in dieser Studie war ebenfalls das Vorhandensein von Aggression. Die möglichen Ergebnisse waren „durchgefallen“, „unsozial“ (Hunde, welche nur den Test auf Verträglichkeit nicht bestanden haben), „grenzwertig“ (Hunde, die milde Aggression in Form von Versteifen oder leichtem Knurren in mindestens einem, aber maximal zwei Testkomponenten gezeigt haben) oder „ohne Problem“ (Referenzkategorie). Die drei Letzteren haben den Test bestanden und wurden zur Adoption freigegeben. Hunde, die den Wesenstest nicht bestanden haben, wurden nicht zur Vermittlung freigegeben. Daher ist auch hier kein Vergleich des Verhaltens dieser Hunde nach Adoption mit dem Verhalten der Hunde, die bestanden haben, möglich.

MCGUIRE (2019) hat sich in ihrer Studie auf das Problemverhalten Ressourcenverteidigung konzentriert (von Futter, Spielzeug, Schlafplätzen) mit Fokus auf die demographischen Daten der Hunde. Sie hat über fünf Jahre eine Studie über 1015 Tierheimhunde in New York betrieben und festgestellt, dass 15% der Hunde in Wesenstests als Ressourcenverteidiger identifiziert wurden. Der Wesenstest nach *Assess-A-Pet*, modifiziert nach BOLLEN et HOROWITZ, erfolgte im Anschluss an die Eingangsuntersuchung innerhalb von drei Tagen nach Aufnahme im Tierheim. Die Verhaltensreaktion im Test mit der Futterschüssel wurde notiert und umfasst sieben Levels, im Test mit dem Kauartikel fünf Levels, je nach Stärke der Reaktion. Eine Klassifikation als Ressourcenverteidiger erfolgte, wenn der Hund entweder Erstarren, „Whale eye“, Knurren, Attackieren, Schnappen und/oder Beißen der *Fake-Hand* zeigt. MCGUIRE (2019) hat Attackieren, Schnappen und Beißen als hochgradiges Verteidigen von Ressourcen eingestuft. Von all den Hunden, die Ressourcenverteidigung zeigten, zeigten 17% hochgradige Ressourcenverteidigung. Meistens wurde jedoch gering- bis mittelgrade Ressourcenverteidigung gezeigt.

In der Studie von MCGUIRE et al. (2020) wurden die Ergebnisse des Wesenstest bezüglich Ressourcenverteidigung mit den Angaben der abgebenden BesitzerInnen und denen der NeubesitzerInnen verglichen. Die Durchführung und Beurteilung des Wesenstests erfolgte analog zu MCGUIRE et al. (2019).

Somit beläuft sich die Messung der diagnostischen Testkriterien für *Assess-A-Pet* lediglich auf die Untersuchung der prognostischen Validität (CHRISTENSEN et al., 2007; BOLLEN et HOROWITZ, 2008; MCGUIRE et al., 2020).

Tab. 20: Messung der diagnostischen Testkriterien im Wesenstest nach Assess-A-Pet

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
Assess-A-Pet und Modifikationen	(✓)	-	-	-	-	-	-	✓

### 3.3.1.2.2 SAFER™

Die initiale Studie (WEISS et GREENBERG, 1997) untersuchte die Auswahl und Akquise geeigneter Hunde für die Assistenzhundebildung aus Tierheimen. Üblicherweise werden Welpen als Assistenzhunde ausgebildet, davon ca. 50% erfolgreich. Auch das Testverfahren war bisher auf die Auswahl und Reevaluierung geeigneter Welpen ausgerichtet. Mittlerweile greifen manche TrainerInnen aber auch auf Hunde aus Tierheimen zurück. In der o.g. Studie wurden neun zufällige Hunde im Alter von zehn Monaten bis zwei Jahren entsprechender Rasse (Labrador Retriever-, Golden Retriever-, Schäfer- und Hütehundmischlinge) einem elfteiligen „Auswahltest“ mit fünf Messparametern unterzogen und anschließend in Gehorsam und Apportieren ausgebildet. Assistenzhunde können verschiedene Aufgaben haben und damit werden auch verschiedene Anforderungen an sie gestellt. Phase 1 dieser Studie bestand aus einem Auswahltest, welcher im ursprünglichen Tierheim stattfand. In Phase 2 und 3 wurden die neun Hunde einem Gehorsamkeitstraining und einem Apportiertraining in einer anderen tierheimähnlichen Unterbringung unterzogen. Danach wurden die Hunde erneut getestet (Phase 4), für zwei Minuten in Gehorsam und anschließend hatten sie in fünf verschiedenen Situationen zu apportieren. Auch dieser Test wurde auf Video aufgezeichnet. Es wurde weiters die Anzahl von notwendigen Korrekturen gezählt. Zur Evaluierung der Test-Retest-Reliabilität ist diese Studie ungeeignet, da zwei verschiedene Testverfahren angewandt wurden.

Nach den unbefriedigenden Ergebnissen der Studie von 1997 hat WEISS im Jahr 2002 einen neuen Auswahltest entwickelt. In dieser Studie wurden 75 Hunde im Alter von sechs Monaten bis zwei Jahren im Tierheim getestet. Es handelte sich um verschiedene Rassen und Mischlinge mit einem Körpergewicht zwischen 22 und 50 Kilogramm. Der Ablauf war ähnlich wie bei WEISS et GREENBERG (1997) in vier Phasen. Der Auswahltest wurde auf 14 Einzeltests erweitert, welche in zufälliger Reihenfolge stattfanden. Vor der Reevaluierung in Phase 4 hatten die TrainerInnen einen Fragebogen auszufüllen.

Die ASPCA® (*American Society for the Prevention of Cruelty to Animals*) hat in weiterer Folge im Jahre 2007 das Programm SAFER™ (*Safety Assessment for Evaluating Rehoming*) erarbeitet. Die Grundlage bildete das von Ph.D. EMILY WEISS entwickelte Instrument zur Aggressionsbewertung. Die beurteilten Hunde sollen in drei Gruppen unterschieden werden können: „*bereit um in die Vermittlung zu gehen*“, „*profitierend von Verhaltensmodifikation*“ oder „*vorsichtiges Management bei Vermittlung notwendig*“. Der Wesenstest soll kurz, exakt und kostengünstig sein. Er dauert inklusive Vorbereitung nur maximal zehn Minuten und inkludiert sieben Einzeltests (Blickkontakt und leichte Bedrängnis, Empfindlichkeit, Reaktion auf Bewegung, Stimme und Anstupsen, Manipulation, Futterverhalten, Spielverhalten und das Verhalten gegenüber anderen Hunden). Zur Verhaltensmodifikation sind vier unterschiedlich strenge Programme vorgesehen. Der finale SAFER™ Test ist auf die Detektion von Aggression inklusive Ressourcenverteidigung spezialisiert und findet zwei bis drei Tage nach Aufnahme im Tierheim statt.

MOHAN-GIBBONS et al. (2012) beschäftigten sich mit dem Verteidigen von Futter. Die Hypothese lautete, dass manche Hunde, die ihr Futter verteidigen, adoptiert werden können ohne dieses Verhalten im neuen Zuhause zu zeigen. Sechshundneunzig (von 6.603) Hunden wurden im SAFER™-Wesenstest als Futterverteidiger identifiziert und zeigten allerdings in den anderen Einzeltests keine Anzeichen von Aggression. Futterverteidiger im SAFER™-Wesenstest sind Hunde, welche in Bezug auf Futter Versteifen, Hinunterschlingen, Erstarren, Knurren und/ Beißen der *Fake-Hand* zeigen. In dieser Studie wurde nur das Verteidigen von Futter, nicht das von Spielzeug, betrachtet. Exkludiert wurden u.a. Hunde der Rassen Rottweiler und Pit Bull, sowie Hunde, die aus Sicherheitsgründen ausgeschlossen werden mussten (weil sie beispielsweise der *Fake-Hand* unter mehrmaligem Zubeißen beim Wegziehen gefolgt sind).

BENNETT et al. (2015) haben untersucht, welchen Unterschied es macht, wenn Hunde direkt nach Eingang im Tierheim oder erst nach drei Tagen Akklimatisierung evaluiert werden, da der Einfluss von akutem und chronischem Stress auf das Zeigen von Aggression noch kontrovers diskutiert wird. Untersucht wurden 33 Hunde durch den *SAFER<sup>TM</sup>*-Test. Diese Studie untersuchte somit die Test-Retest-Reliabilität. Die AutorInnen der Studie empfahlen, den Wesenstest eventuell monatlich durchzuführen, abhängig unter anderem von den Möglichkeiten der Verhaltensmodifikation vor Ort.

MOHAN-GIBBONS et al. (2018) untersuchten den Einfluss der Eliminierung des Einzeltests „*Futterverteidigung*“ auf das Ergebnis der *SAFER<sup>TM</sup>*-Wesenstests. Für zwei Monate wurden Hunde in neun Tierheimen nach dem „normalen“ Eingangstest evaluiert. Sieben der neun Tierheime verwendeten den *SAFER<sup>TM</sup>* Test, die anderen beiden verwendeten einen eigenen standardisierten Wesenstest. Neben der Identifikation von Futterverteidigern im Wesenstest wurden auch jene Hunde als futterverteidigend identifiziert, sofern die VorbesitzerInnen dies berichteten oder das Personal im Tierheim ein solches Verhalten beobachtet hatte. In den anschließenden zwei Monaten wurden die Hunde bei Aufnahme im Tierheim zwar weiterhin evaluiert, allerdings wurde der Einzeltest für Futterverteidigung ausgelassen. Berichte durch VorbesitzerInnen oder Personal wurden weiterhin wahrgenommen. Für beide Testphasen wurden monatlich folgende Daten erhoben und verglichen: Gesamtzahl der aufgenommenen Hunde; Gesamtzahl der zurückgegebenen Hunde nach Adoption innerhalb von 30 Tagen; durchschnittliche Aufenthaltsdauer im Tierheim; Gesamtzahl der adoptierten, überstellten und euthanasierten Hunde; Gesamtzahl an Bissen/ Verletzungen sowohl während des Aufenthalts im Tierheim, als auch nach der Adoption (sofern berichtet).

Somit wurden für den *SAFER<sup>TM</sup>*-Wesenstest die diagnostischen Kriterien Test-Retest-Reliabilität (BENNETT et al., 2015), Inter-Rater-Reliabilität (WEISS, 2002) und prognostische Validität (MOHAN-GIBBONS et al., 2012) untersucht, siehe Tab. 21.

Tab. 21: Messung der diagnostischen Testkriterien im Wesenstest nach SAFER™

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
SAFER™ und Modifikationen®	(✓)	-	-	✓	✓	-	-	✓

### 3.3.1.2.3 Evaluation of fearful dogs der ASPCA®

In ihrer Abschlussarbeit untersuchte MATTERS (2016) die Intra-Rater-Reliabilität und Inter-Rater-Reliabilität von vier erfahrenen Beurteilerinnen bei der Verwendung des ASPCA® *Behavioral Rehabilitation Center Behavior Evaluation of fearful dogs* in zehn Einzeltests und in 29 verschiedenen Situationen. Dieser Test weist Teilinhalte von SAFER™, Assess-A-Pet und C-BARQ auf, aber auch einzigartige Einzeltests.

Tab. 22: Messung der diagnostischen Testkriterien im Wesenstest der ASPCA®

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
<i>Evaluation of fearful dogs ASPCA®</i>	(✓)	-	-	-	✓	✓	-	-



#### 3.3.1.2.4 **Match-Up und Modifikationen**

Das *Match-Up II*® *Shelter Dog Rehoming Program* wurde von der *Animal Rescue League of Boston* entwickelt und besteht aus fünf Teilen:

1. der Verhaltensgeschichte des Hundes (wird durch einen Fragebogen bei Aufnahme im Tierheim erfasst)
2. einem Wesenstest
3. einer Auswertung der Persönlichkeit (in den Eigenschaften *Freundlichkeit, Angst, Aufregung, Aggressivität, Verspieltheit* und *Intelligenz*),
4. dem Verhalten des Hundes im Tierheim
5. einer „Verhaltens-Triage“ (zusammengesetzt aus den Punkten 1,2 und 4).

Der Wesenstest findet erstmals etwa zwei bis drei Tage nach Aufnahme im Tierheim statt (bei sehr ängstlichen Hunden bis zu einer Woche später) und wird monatlich wiederholt, sowie bei jeder signifikanten Verhaltensänderung und bei Rückgabe des Hundes in das Tierheim nach Vergabe. Der Test beinhaltet elf Teiltests zu Leinenführigkeit, Verhalten im Raum, Kommandos, Manipulation, Spielverhalten, Reaktion auf Bewegung und Lärm, Reaktion auf eine „*Fake-Hand*“, besitzergreifendem Verhalten, Reaktion auf eine Puppe und eine „eigenartige“ fremde Person, sowie zur Verträglichkeit mit anderen Hunden. Die Ergebnisse der Verhaltens-Triage entscheiden über das weitere Vorgehen.

In der Studie von DOWLING-GUYER et al. (2011) wurde die Fähigkeit des *Match-Up*-Wesenstests untersucht, jene Verhaltensmuster zu erkennen, die Teile von der Persönlichkeit des Hundes widerspiegeln sollen. Hierfür wurden die Ergebnisse von 668 Wesenstests von Tierheimhunden ausgewertet. Der verwendete Wesenstest bestand aus 19 Einzeltests, von denen ein Einzeltest exkludiert wurde, da er anders aufgeschlüsselt wurde als die anderen. Der Inhalt der Studie ist die Untersuchung der internen Konsistenz für die vier Messparameter. Die Messparameter wurden in der Studie mit den Persönlichkeitsmerkmalen des Hundes gleichgesetzt und somit wurde die Validität in Abhängigkeit der Situation und Voraussagekraft nicht weiter in Frage gestellt. In der Studie wurde angemerkt, dass eine Untersuchung der Inter-Rater-Reliabilität aufgrund fehlender Informationen nicht möglich war. Hunde, die zu gefährlich zum Evaluieren waren, wurden nicht einbezogen, wodurch lediglich eine Hundepopulation ohne erhebliche Verhaltensprobleme untersucht wurde.

Die Studie von MARDER et al. (2013) untersuchte retrospektiv das Auftreten von Futteraggression im Wesenstest und Futteraggression nach der Adoption, berichtet durch die NeubesitzerInnen. Im Zuge des *Match-Up* Wesenstests wurden 97 Hunde nach dem Auftreten von Futteraggression klassifiziert. Bei 20.6% (20 Hunde) der Hunde wurde das Vorhandensein von Futteraggression festgestellt. Exkludiert wurden nur Hunde, die zu gefährlich waren, um gehandelt zu werden. In dieser Studie wurden nur jene beiden Einzeltests betrachtet, die mit Futteraggression in Verbindung stehen. Knurren, Zähne zeigen, schnappen, attackieren oder beißen in mindestens einem Einzeltest wurde als Vorhandensein von Futteraggression gewertet. Keiner der Hunde wurde während des Tierheimaufenthaltes bezüglich der Futteraggression trainiert. Bei der Vergabe von positiv getesteten Hunden wurden die NeubesitzerInnen genau über das Management während des Fressens instruiert.

Die Studie von SHABELANSKY et al. (2015) untersuchte die Konsistenz der Reaktionen von Hunden im Wesenstest auf echte und unechte Stimulushunde, was als Untersuchung der Objektivität verstanden werden kann.

Somit wurden für den *Match-Up*-Wesenstest die interne Konsistenz (DOWLING-GUYER et al., 2011), sowie die prognostische Validität (MARDER et al., 2013) untersucht.

Tab. 23: Messung der diagnostischen Testkriterien im Wesenstest nach Match-Up

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
Match-Up und Modifikationen	✓	-	✓	-	-	-	-	✓

### **3.3.1.3 Wesenstests in Australien**

#### **3.3.1.3.1 Wesenstest der RSPCA**

Die Hunde im Wesenstest der *RSPCA* werden nicht direkt nach der Aufnahme im Tierheim getestet, sondern erst, bevor sie in die Vermittlung gehen. Dieser Test wird von der *Royal Society for the Protection of Cruelty to Animals (RSPCA)* in Australien verwendet. Der Test beinhaltet Berührung und Manipulation durch eine fremde Person, eine tierärztliche Untersuchung, Bedrängung, den Hund auf den Rücken rollen, das Wegnehmen von Futter, Reaktionen auf Lärm und Bewegung, Spielverhalten, Alleinlassen und Verträglichkeit mit anderen Hunden. Der Punktescore in der Auswertung entscheidet über die weitere Vorgehensweise mit dem getesteten Hund und unterscheidet zwischen „*Test bestanden*“, „*Verhaltenstherapie und Testwiederholung nach einer Woche*“ und „*Test nicht bestanden, Euthanasie*“, allerdings fließen Erfahrungen des Tierheimpersonals mit dem jeweiligen Hund ebenso in die Entscheidung ein.

In der Studie von POULSEN et al. (2010) wurden 236 Hunde getestet. Einhundertdreißig Hunde haben den Test bestanden, 24 Hunde wurden für Verhaltensmodifikationsprogramme vorgeschlagen (und erneuter Test nach einer Woche) und 82 Hunde haben nicht bestanden (Euthanasie). Die Hauptgründe für Nicht-Bestehen waren generelle Angst und Aggression gegenüber anderen Hunden. Es wurde die Test-Retest-Reliabilität gemessen, nach der Vermittlung wurden außerdem die NeubesitzerInnen befragt.

Durch CLAY et al. (2020a) wurden 123 Hunde im Tierheim einem Wesenstest unterzogen und einen Monat nach Vermittlung wurden die NeubesitzerInnen via Umfrage zum Verhalten des Hundes im neuen Zuhause befragt. Der Wesenstest erfolgte fünf Tage nach Aufnahme im Tierheim durch den Wesenstest der *RSPCA*.

In der Studie von CLAY et al. (2020b) wurden die Ergebnisse der BesitzerInnenbefragung durch *C-BARQ* verglichen mit den Ergebnissen des Wesenstests der *RSPCA* während eines Besuches im Tierheim. Einhundertsieben Hunde wurden somit einmal durch ihre BesitzerInnen und einmal durch die AutorInnen beurteilt, ein bis zwei Wochen später. Der Wesenstest fand in Abwesenheit der BesitzerInnen statt. Die getestete Population spiegelt dennoch keine Hundepopulation im Tierheim wider, sondern besteht aus Hunden in Privatbesitz. Nur 44.9% der Hunde waren ursprünglich aus dem Tierheim.

Somit wurden für den Wesenstest der RSPCA die diagnostischen Testkriterien Test-Retest-Reliabilität (POULSEN et al., 2010) und prognostische Validität (POULSEN et al., 2010; CLAY et al., 2020a) untersucht.

Tab. 24: Messung der diagnostischen Testkriterien im Wesenstest der RSPCA

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
RSPCA	(✓)	-	-	✓	-	-	-	✓

### 3.3.1.3.2 B.A.R.K.

Das Ziel der Studie von MORNEMENT et al. (2014) war die Untersuchung des *B.A.R.K.*-Wesenstests hinsichtlich seiner Reliabilität und Voraussagekraft. *B.A.R.K.* steht für *behavioural assessment for re-homing K9's*. Der Test besteht aus zwei Teilen, Teil A als Basistest und Teil B als Zusatztest für jene Hunde, die Teil A erfolgreich und ohne Zeichen von Aggression absolviert haben. Ein niedriger Score sprach für „*erwünschtes Verhalten*“ (beispielsweise *nicht nervös/ sehr freundlich*) und ein hoher Score für „*unerwünschtes Verhalten*“ (beispielsweise *extrem ängstlich/ sehr unfreundlich*). MORNEMENT et al. haben die Test-Retest-Reliabilität, die Inter-Rater-Reliabilität und die Vorhersagekraft gemessen. Insgesamt wurden (inklusive der 48 Hunde zur Testung der Test-Retest- und Inter-Rater-Reliabilität) 102 Hunde getestet. Hunde, die im Tierheim Aggression oder extreme Angst gezeigt hatten, wurden allerdings nicht in die Studie inkludiert.

In einer Folgestudie untersuchen MORNEMENT et al. (2015) die Fähigkeit des *B.A.R.K.*-Wesenstests Verhaltenstendenzen im neuen Zuhause vorausszusagen. Vierundsiebzig Hunde wurden im *B.A.R.K.*-Wesenstest beurteilt. Dieser fand analog zu MORNEMENT et al.

(2014) statt und wurde erst nach mindestens drei Tagen Akklimatisierung im Tierheim durchgeführt. Alle Hunde hatten zuvor bereits den tierheimeigenen Wesenstest bestanden.

Tab. 25: Messung der diagnostischen Testkriterien im Wesenstest nach B.A.R.K.

Wesenstest	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität	prognostische Validität
B.A.R.K.	(✓)	-	-	✓	✓	-	-	✓

### 3.3.2 Vergleichende Übersicht aller Wesenstests

In der folgenden Tabelle (Tab. 26) findet sich eine Übersicht über die gemessenen diagnostischen Kriterien, aufgeschlüsselt auf die einzelnen Studien. Leider misst keine der Studien alle relevanten Kriterien, ebenfalls wurden für keinen Wesenstest all diese Kriterien gemessen (siehe Tabellen 12 bis 24).

In Tabelle 27 findet sich eine Übersicht, welche Messparameter in den Wesenstests berücksichtigt wurden.

Tab. 26: Übersicht über die untersuchten diagnostischen Kriterien in den Studien

Wesenstest	Studie	Anzahl Messparameter	Anzahl Einzeltests	Objektivität			Reliabilität			Validität	
				standardisierter Ablauf	Anzahl BeurteilerInnen	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität
	Van der Borg et al., 1991	5	21	✓	3						✓
Assess-A-Pet	Christensen et al., 2007	7	14	✓	1						✓
	Bollen et Horowitz, 2008	1	9	✓	1						✓
	McGuire, 2019	1	9	✓	2						
	McGuire et al., 2020	1	9	✓	2						✓
SAFER™-Test	Weiss et Greenberg, 1997	4	11	✓	3						
	Weiss, 2002	5	14	✓	2			✓			
	Mohan-Gibbons et al., 2012	1	7	✓	2						✓
	Bennett et al., 2015	1	7	✓	2			✓			
	Mohan-Gibbons et al., 2018	1	7	✓	2						
RSPCA	Poulsen et al., 2010	3	28	✓	2			✓			✓
	Clay et al., 2020 (a)	17	11	✓	2						✓
	Clay et al., 2020 (b)	5	9	✓	1						

TT	Valsecchi et al., 2011	7	22	✓	2		✓	✓	✓	✓	✓	
	Barnard et al., 2019	7	22	✓	2	✓	✓				✓	
B.A.R.K.	Mornement et al., 2014	5	12	✓	2			✓	✓			✓
	Mornement et al., 2015	5	12	✓	2							✓
	Zilocchi et al., 2015	3	13	✓	2							✓
	Matters, 2016	3	10	✓	2				✓	✓		
	Frank, 2007	2	12	✓	1			✓				✓
	Kis et al., 2014	3	3	✓	2			✓	✓			
	Diesel et al., 2008	6	3	(✓)	1				✓	✓		
Match-Up	Dowling-Guyer et al., 2011	4	18	✓	2		✓					
	Marder et al., 2013	1	11	✓	2							✓
	Shabelansky et al., 2015	3		✓	2							
Ethotest	Lucidi et al., 2005	3	36	✓	5							
	Sticco et al., 2011	3	16	✓	4							✓

Die gemessenen Parameter variieren stark, obwohl der Ablauf der Wesenstests sich meist in den Grundzügen ähnelt. So wird beispielsweise in nahezu allen Wesenstests die Reaktion des Hundes auf die/den ihm unbekannte/n PrüferIn beurteilt. Dennoch folgen in einigen Wesenstests (z.B. BOLLEN et HOROWITZ, 2008) noch zusätzliche Einzeltests, in denen fremde, teilweise bedrohlich wirkende und/oder handelnde Personen als Stimulus agieren.

Einige AutorInnen legen ihr Hauptaugenmerk auf das Verteidigen von Ressourcen (MCGUIRE, 2019; MCGUIRE et al., 2020; MOHAN-GIBBONS et al., 2012; MOHAN-GIBBONS et al., 2018; BENNETT et al., 2015; MARDER et al., 2013). Manche Wesenstests haben eine Beurteilung der Vermittelbarkeit des Hundes als Ziel (FRANK, 2007; ZILOCCHI et al., 2015; LUCIDI et al., 2005, *Ethotest*; STICCO et al., 2011, *Ethotest*).

Tab. 27: Messparameter, welche in den Wesenstests berücksichtigt wurden

Studie	Messparameter												
	Aggression ggü. Menschen	Angst/ Ängstlichkeit	Artgenossen	Trennungsangst	(Bemühen um) Aufmerksamkeit	Erregung/ Nervösität	Trainierbarkeit/ Gehorsam	Ziehen an der Leine	Bellen	Ressourcen-verteidigung	Freundlichkeit/ Umgänglichkeit	Vermittelbarkeit	weitere
Van der Borg et al., 1991	+	+	Aggression	+			+	+	+	+			+
Christensen et al., 2007 (Assess-A-Pet)	Beute/ Kinder BesitzerIn/ territorial/	Angst- aggression	Aggression							+			
Bollen et Horowitz, 2008 (Assess-A-Pet)	+												
McGuire, 2019 (Assess-A-Pet)										Futter/ Spielzeug/ Schlafplatz			
McGuire et al., 2020 (Assess-A-Pet)										Futter/ Spielzeug			
Weiss et Greenberg, 1997 (SAFER™)		+			+	+	+						+





Dowling-Guyer et al., 2011 (Match Up)	+	+			+						+		+
Marder et al., 2013 (Match Up)										Futter			
Shabelansky et al., 2015 (Match Up)			Freundlichkeit/ Angst/ Aggression										
Kis et al., 2014	+	+											
Frank 2007	+	+			+	+						+	+
Zilocchi et al., 2015	+	+									+	+	
Lucidi et al., 2005 (Ethotest©)	+						+				+	+	
Sticco et al., 2011 (Ethotest©)	+	+					+				+	+	

### 3.3.3 Diagnostische Testkriterien der Wesenstests

#### 3.3.3.1 Objektivität von Wesenstests

Alle untersuchten Wesenstests haben einen standardisierten Ablauf. Die Testinstruktionen wurden exakt und schriftlich festgehalten und die möglichen Reaktionen des Hundes wurden genau aufgeschlüsselt, um Interpretationen zu vermeiden. Lediglich der Wesenstest, welcher von der britischen Tierschutzorganisation *Dogs Trust* verwendet wird, war (noch) nicht standardisiert zwischen den verschiedenen Tierheimen (DIESEL et al., 2008). Für die Videoaufnahmen zur Untersuchung der Intra- und Inter-Rater-Reliabilität in der Studie wurde der Testablauf aber für die teilnehmenden Hunde weitestgehend standardisiert. Die Aufzeichnung des Wesenstests in Form einer Videoaufnahme ist ein maßgeblicher Parameter für Objektivität, da somit der Wesenstest jederzeit reevaluiert werden kann und Verhaltensweisen, welche in Echtzeit eventuell nicht erfasst werden, beurteilt werden können. In einigen Studien wurde dies umgesetzt (siehe Tab. 28).

In kaum einer Studie wurde jedoch explizit die Praktikabilität untersucht. Die Praktikabilität sinkt mit der Anzahl der Einzeltests (FRANK, 2007). Mit steigender Anzahl der Einzeltests steigt dagegen die Wahrscheinlichkeit, Aggression oder anderes Problemverhalten zu detektieren. Von großer Bedeutung ist hierbei auch die geplante Dauer des Wesenstests. Diese reichte von drei Minuten (Wesenstest nach KLAUSZ et al.) bis hin zu 90 Minuten (Wesenstest nach VAN DER BORG et al.). Je länger und komplexer ein Wesenstest aufgebaut ist, desto mangelhafter sind Objektivität (durch die steigende Wahrscheinlichkeit von Ablenkungen und Abweichungen) und Praktikabilität.

Weiters haben wenige AutorInnen die Replizierbarkeit von Wesenstests gemessen. Je replizierbarer ein Wesenstest ist, desto größer ist dessen Objektivität. Ein wichtiger Punkt hierbei ist die Verwendung von sogenannten *Fakehunden* oder kindergroßen Puppen zur Testung der Verträglichkeit und Kinderfreundlichkeit. Durch diese Hilfsmittel ist ein Wesenstest besser replizierbar als durch den Einsatz lebendiger Stimuli. BARNARD et al. (2019, *Temperament Test*) haben beispielsweise die Reaktionen von Hunden auf einen *Fakehund* und einen lebenden Hund verglichen und das Vorhandensein von Konsistenz im Verhalten untersucht. Diese Fragestellung stellt sich besonders aus Gründen der Sicherheit für den Stimulushund und die PrüferInnen. Es wurden jeweils zwei echte Hunde in zwei

verschiedenen Größen und analog zwei unechte Hunde verwendet. Die echten Hunde wurden nach guter Verträglichkeit mit anderen Hunden ausgewählt. Jeder Stimulushund wurde dem zu testenden Hund in randomisierter Reihenfolge in drei Phasen vorgeführt (Distanz, Annäherung, Nähe), zwischen den Stimulushunden wurde für einige Minuten spazieren gegangen. Die Reaktionen wurden in einem Ethogramm beurteilt und den Messparametern *Selbstvertrauen*, *Vorsicht*, *Bestimmtheit* und *Angst* zugeordnet. Die Hypothese war, dass die Reaktion des Hundes auf den Stimulushund abhängig ist von dessen Größe. Die Reaktionen des Hundes auf den echten vs. den unechten Stimulushund waren signifikant korreliert und waren unabhängig von dessen Größe. Somit wurde der *Temperament Test* als leicht replizierbar und konsistent in seinen Grundergebnissen beschrieben. Auch SHABELANSKY et al. (2015; *Match-Up-Wesenstest*) untersuchten die Konsistenz der Reaktionen von Hunden im Wesenstest auf echte und unechte Stimulushunde. Untersucht wurden 45 Hunde unter zwei verschiedenen Konditionen – einem Plüschhund und einem lebendigen Hund. Es wurde für alle getesteten Hunde jeweils der selbe Hund verwendet. Die Reaktionen wurden standardisiert beurteilt nach *Anwesenheit/Abwesenheit* und in *Freundlichkeit*, *Aggression* und *Angst* eingeteilt. Eine knapp gute Übereinstimmung erreichte *Freundlichkeit* gegenüber dem Plüsch- und dem lebendigen Hund. Die Übereinstimmung für *Angst* war nur moderat und für *Aggression* war keine Übereinstimmung vorhanden. Die Sinnhaftigkeit der Verwendung eines Plüschhundes statt einem lebendigen Hund ist daher sehr fraglich, da *Aggression* wohl den wichtigsten Messparameter für die meisten zukünftigen BesitzerInnen darstellen könnte. Es ist außerdem unzureichend untersucht, inwiefern der Stimulushund durch sein Verhalten die Ergebnisse des Einzeltests beeinflussen kann.

Unterschiedliche BeurteilerInnen und AssistentInnen haben einen Einfluss auf die Standardisation von Wesenstests (TAYLOR et MILLS, 2016). Die Ergebnisse eines Wesenstests könnten beispielsweise zwischen einem männlichen und einer weiblichen BeurteilerIn eine nur schwache Übereinstimmung erreichen. Einen ebenso großen Unterschied könnte es machen, ob die/der PrüferIn dem Hund bekannt ist oder nicht. In einigen Wesenstests wurde das Geschlecht der prüfenden Person daher vorgeschrieben und in den meisten Wesenstestinstruktionen wurde vermerkt, dass die/der PrüferIn dem Hund nicht vertraut sein darf. Neben den genannten Aspekten spielen auch andere Faktoren eine Rolle, wie beispielsweise die Körpersprache und das Verhalten gegenüber dem Hund. Die Prüferin/ der

Prüfer sollte daher genauestens instruiert werden und maximale Kompetenz auf dem Gebiet des Hundeverhaltens aufweisen. Nur so ist eine objektive Beurteilung gewährleistet. Aus diesem Grund wurde das Häkchen in allen Tabellen der einzelnen Wesenstests lediglich in Klammern gesetzt, da die BeurteilerInnen/TesterInnen sich in den Wesenstests deutlich unterscheiden.

Die Wesenstests *nach FRANK (2007)* und *nach DIESEL et al. (2008)* wurden durch eine (einzelne) Testerin durchgeführt. Die TeilnehmerInnen wurden vorher ausgiebig geschult über Hundeverhalten, weswegen das Wissen über die Terminologien der einzelnen Verhaltensweisen vorausgesetzt wurde. Im Gegensatz zu FRANK haben DIESEL et al. die Wesenstests auf Video aufgenommen (siehe Tab. 28). In den meisten Wesenstests sind mindestens zwei PrüferInnen vorgeschrieben. Der Test *nach VAN DER BORG et al. (1991)* wurde durch drei Personen durchgeführt: einer handelnden Prüferin, einer assistierenden Prüferin und einem assistierendem Prüfer. Die Reaktionen wurden auf einem speziellen Formblatt direkt vor Ort, und wenn nötig später von den Videoaufnahmen aufgezeichnet.

Der *Assess-A-Pet* Wesenstest wird in ruhiger Umgebung durchgeführt durch zwei bis drei, dem Hund nicht bekannte Personen. *Assess-A-Pet* soll höflich, nicht-konfrontativ und nicht-bedrohlich sein und somit den Hund neutral und objektiv bewerten können. Die PrüferInnen haben erfahrene Personen zu sein und müssen idealerweise zu zweit oder zu dritt sein, da nicht eine Person allein testen und beobachten kann. Die/der PrüferIn darf sich dem Hund gegenüber nicht als professionell zu erkennen geben, da davon auszugehen ist, dass die/der typische NeubesitzerIn sich ebenfalls nicht professionell verhält im Umgang mit dem Hund und die Reaktionen des Hundes dadurch verfälscht werden könnten. Die modifizierten Tests nach STERNBERG durch CHRISTENSEN et al. (2007) im Tierheim wurden allerdings durch nur eine/n einzelnen gut ausgebildete/n TierheimmitarbeiterIn durchgeführt. Der Wesenstest nach *Assess-A-Pet*, modifiziert nach BOLLEN et HOROWITZ, in der Studie von MCGUIRE (2019) wurde durch zwei TierheimmitarbeiterInnen durchgeführt (eine Person beurteilt, eine Person schreibt).

Im Wesenstest nach *SAFER™* sind ein/e GutachterIn und ein/e BeobachterIn notwendig. Die Personen bedürfen eines individuellen Zertifikates. Dieses Zertifikat ist begrenzt gültig und kann nur erlangt werden, indem drei Videos und entsprechende Arbeitsblätter zur Beurteilung eingesandt werden. Die/Der PrüferIn im Auswahltest in der initialen Studie von

WEISS et GREENBERG (1997) war erfahren in der Arbeit mit Hunden und im Umgang ähnlich wie AssistenzhundetrainerInnen. Bei WEISS (2002) gab es vier PrüferInnen, alle mit mindestens fünf Jahren professioneller Erfahrung. Beide Wesenstests (Test und Wiederholungstest) in BENNETT et al. (2015) wurden durch jeweils dieselben zertifizierten PrüferInnen durchgeführt, welche außer in den beiden Tests keinen Kontakt zu den Hunden hatten.

Die Tests der *RSPCA* durch POULSEN et al. (2010) wurden durch jeweils zwei von vier verfügbaren PrüferInnen durchgeführt und beobachtet durch POULSEN, solange bis keine Differenzen im Scoring zwischen den PrüferInnen und POULSEN mehr vorhanden waren – sodass POULSEN den Wesenstest in gleicher Art und Weise allein durchführen kann. CLAY et al. (2020a) haben den Wesenstest (der *RSPCA*) durch eine/n PrüferIn und eine/n BeobachterIn durchgeführt. Der Wesenstest in CLAY et al. (2020b) wurde durch eine/n der AutorInnen durchgeführt.

Im *Temperament-Test* (VALSECCHI et al., 2011) ist das Verhalten der prüfenden Personen besonders genau aufgeschlüsselt. Diese müssen professionellen Umgang mit Hunden aufweisen (z.B. TierärztInnen, Tierheimpersonal). Zur zusätzlichen Vereinheitlichung wurden für jede/n TesterIn Übungsvideos von drei Evaluierungen aufgenommen und bezüglich Körperhaltung der TesterIn/ des Testers und Auswertung des Verhaltens des Hundes diskutiert. Auch die Tageszeit der Wesenstests wurde vereinheitlicht. Bei jedem Wesenstest prüften ein/e TesterIn und ein/e AssistentIn. Die/der PrüferIn muss dem Hund fremd sein. Um die Objektivität explizit messen zu können, wurde eine Evaluierung der Testkriterien mithilfe einer unabhängigen Hundexpertin/ eines unabhängigen Hundexperten (HundeverhaltenswissenschaftlerIn und zertifizierte/r HundetrainerIn) durchgeführt. Diese ergab für die meisten Testkriterien erhebliche bis perfekte Übereinstimmung zwischen TesterIn A/B und HundexpertIn, aber für *Umgänglichkeit mit Artgenossen* und *Umgänglichkeit mit Menschen* war die Übereinstimmung lediglich ausreichend bis moderat. Zwischen Tester B und der Expertin /dem Experten gab es keine signifikante Korrelation bezüglich *Umgänglichkeit mit Menschen*. Die PrüferInnen in der Studie von BARNARD et al. (2019) waren erfahren im Umgang mit Hunden und kannten die zu testenden Hunde nicht. Die Prüferin (weiblich) nahm den Test ab, der Prüfer (männlich) beobachtete und beurteilte die Reaktionen des Hundes. Eine initiale Trainingsphase der PrüferInnen ging dem voraus. Die Praktikabilität wird hier definiert dadurch, wie einfach der Wesenstest und die Beurteilung des Verhaltens

durchgeführt werden kann, als auch dadurch, wie durchführbar er in unterschiedlichem Umfeld ist. Ein Problem stellt hierbei die Verfügbarkeit eines angemessenen Stimulushundes dar, welcher ruhig und gut sozialisiert sein sollte, siehe „*Praktikabilität*“.

Im B.A.R.K. Wesenstest (MORNEMENT et al., 2014) hat eine Person den Hund durch den Wesenstest geleitet und eine zweite Person hat beobachtet und beurteilt. Die Personen wurden durch Workshops vorbereitet und ihre Kompetenz wurde durch mehrere Probedurchläufe getestet. Die ausreichende Untersuchung der Praktikabilität (*feasibility*) des Tests durch eine Befragung der TierheimmitarbeiterInnen war aufgrund der niedrigen Teilnahme (verursacht durch Personalwechsel und Naturkatastrophen) nicht möglich.

Der Wesenstest in der Beurteilung von ZILOCCHI et al. (2015) wurde durch zwei Personen durchgeführt (VerhaltensmedizinerIn und HundetrainerIn) und anschließend wurden die Ergebnisse verglichen. Auch der Wesenstest *nach* KLAUSZ et al. wurde ebenfalls durch zwei PrüferInnen durchgeführt. Jeweils zwei PrüferInnen führten auch den *Match-Up* Wesenstest in der Studie von DOWLING-GUYER et al. (2011) durch. Leider hatten diese teilweise vor dem Wesenstest bereits Kontakt mit den zu testenden Hunden. Auch das Geschlecht der PrüferInnen wurde nicht standardisiert. Zwei erfahrene und ausgebildete Personen haben den *Match-Up* Wesenstest bei MARDER et al. (2013) durchgeführt.

Die Beobachtungsteams im *Ethotest* (LUCIDI et al., 2005) bestanden aus einer/einem GutachterIn, zwei InstruktorInnen und zwei BeobachterInnen. Drei von vier Testteilen bei STICCO et al. (2011) wurden durch zwei PrüferInnen durchgeführt, von denen eine/r dem Hund bekannt und eine/r dem Hund fremd war (und daher von Test zu Test wechseln musste). Die Reaktionen wurden für die beiden Personen in der Bewertung unterschiedlich gewichtet.

Tab. 28: erfüllte Kriterien zu Praktikabilität und Objektivität

Wesenstest	Testdauer	Videoaufnahme	Anzahl BeurteilerInnen
nach VAN DER BORG	90 Minuten	✓	3
Assess-A-Pet	20 Minuten	-	1 bis 3

SAFER™	10 Minuten	✓	2 bis 3
nach FRANK	15 Minuten	-	1
RSPCA	15 – 20 Minuten	-	1 bis 2
Temperament test	keine Angabe	✓	2
B.A.R.K.	20 Minuten	-	2
nach ZILOCCHI	keine Angabe	-	2
Match-Up	15 – 30 Minuten	-	2
Dog´s Trust	keine Angabe	✓	1
Evaluation of fearful dogs (ASPCA®)	keine Angabe	✓	2
nach KLAUSZ et al.	3 Minuten	✓	2
Ethotest u. Ethotest-R	keine Angabe	-	4 bis 5

### 3.3.3.2 Reliabilität von Wesenstests

#### 3.3.3.2.1 Test-Retest-Reliabilität von Wesenstests

Um die Test-Retest-Reliabilität zu beurteilen, ist die Wiederholung des Wesenstests unter den gleichen Bedingungen notwendig. Untersucht wurde die Test-Retest-Reliabilität für die Wesenstests nach *SAFER™* (BENNETT et al., 2015), der *RSPCA* (POULSEN et al., 2010), für den *Temperament Test* (VALSECCHI et al., 2011), für den *B.A.R.K.* Wesenstest (MORNEMENT et al., 2014) und für die Wesenstests *nach FRANK* (2007) und *nach KLAUSZ et al.* (KIS et al., 2014). Für den *Assess-A-Pet* Wesenstest, *Match-Up*, *Ethotest* und die weiteren Tests fehlen bisher jegliche Untersuchungen zur Test-Retest-Reliabilität (siehe Tab. 26).



Die Zeitpunkte des initialen Wesenstests nach Aufnahme im Tierheim und des Wiederholungstests variieren von Wesenstest zu Wesenstest (siehe Tab. 29). In manchen Studien wurde kein definierter Zeitpunkt angegeben (VAN DER BORG et al., 1991; MATTERS, 2016; WEISS et GREENBERG, 1997; WEISS, 2002; DIESEL et al., 2008). Teilweise variiert der Testzeitpunkt in der Praxis nach Herkunft des Hundes: während Fundhunde meist erst nach einer Abholfrist<sup>5</sup> evaluiert werden, können Abgabehunde vom ersten Tag an evaluiert werden (MORNEMENT et al., 2010). Der Zeitpunkt der Verhaltensevaluierung könnte eine große Rolle spielen, besonders in Zusammenhang mit dem Stresslevel.

Tab. 29: Übersicht über Zeitpunkt des Wesenstests im Tierheim

Studie/ Wesenstest	Zeitpunkt der Testung	Wiederholungstest
CLAY et al. (2020b)	nach 15 Minuten Akklimatisierung	kein Wiederholungstest
BENNETT et al. (2015)/ SAFER™	1 Tag nach Aufnahme im Tierheim	3 Tage nach Aufnahme im Tierheim
KIS et al. (2015)	1-2 Tage nach Aufnahme im Tierheim	2 Wochen nach dem ersten Test
MOHAN-GIBBONS et al. (2012)/ SAFER™	2-3 Tage nach Aufnahme im Tierheim	kein Wiederholungstest
MCGUIRE (2019) MCGUIRE et al. (2020)/ Assess-A-Pet	innerhalb von 3 Tagen nach Aufnahme im Tierheim	kein Wiederholungstest
MORNEMENT et al. (2014; 2015)/ B.A.R.K.	mindestens 3 Tage nach Aufnahme im Tierheim	24 Stunden nach dem ersten Test
BOLLEN et HOROWITZ, 2008/ Assess-A-Pet	2-4 Tage nach Aufnahme im Tierheim	kein Wiederholungstest
DOWLING-GUYER et al. (2011)/ Match-Up	2-4 Tage nach Aufnahme im Tierheim	kein Wiederholungstest
MARDER et al. (2013)/	2-7 Tage nach Aufnahme im	kein Wiederholungstest

<sup>5</sup> Unter der Abholfrist ist hier eine gewisse Spanne an Tagen oder Wochen gemeint, in denen die Hunde für die etwaige Abholung durch ihre ehemaligen BesitzerInnen zurückgehalten werden.

Match-Up	Tierheim	
CHRISTENSEN et al. (2007)/ Assess-A-Pet	4 Tage nach Aufnahme im Tierheim	kein Wiederholungstest
FRANK (2007)	innerhalb von 5 Tagen nach Aufnahme im Tierheim	4 Wochen nach dem ersten Test
CLAY et al. (2020a)	5 Tage nach Aufnahme im Tierheim	kein Wiederholungstest
BARNARD et al. (2019)	mindestens 1 Woche nach Aufnahme im Tierheim	kein Wiederholungstest
LUCIDI et al. (2005)/ Ethotest	7 Tage bis 2 Jahre nach Aufnahme im Tierheim	kein Wiederholungstest
VALSECCHI et al. (2011)/ Temperament Test	20 Tage (oder 1 Jahr nach Aufnahme im Tierheim)	40 Tage nach dem ersten Test
STICCO et al. (2011)/b Ethotest	2 Monate bis 6 Jahre nach Aufnahme im Tierheim	kein Wiederholungstest

BENNETT et al. (2015) haben die Durchführung des SAFER™ Wesenstests zum Zeitpunkt der Aufnahme im Tierheim mit der nach drei Tagen verglichen. Es konnten nur moderate Übereinstimmungen ohne einheitliches zeitliches Muster gemessen werden. Die Test-Retest-Reliabilität scheint unzureichend zu sein für so gut wie alle Einzeltests. Es stellt sich die Frage, wie der Wesenstest nach einem weiteren Monat ausfallen würde, es wären erneut starke Diskrepanzen zu erwarten. Eine besonders starke Diskrepanz erreichten die Einzeltests „Augenkontakt“, „Futterverhalten“ und der erste „Kneiftest“. Die Übereinstimmung war gering bei den Einzeltests „Augenkontakt“, erster „Kneiftest“ und bei der Verträglichkeit mit anderen Hunden. Eine mittelgradige Übereinstimmung ergaben die Einzeltests „Berührungssensitivität“, „Spielverhalten“ und „Futterverhalten“. Nur bei der Wiederholung des „Kneiftests“ war die Übereinstimmung hoch (siehe Tab. 30). Einige Einzeltests mussten aufgrund von Aggression in vorherigen Einzeltests übersprungen werden, was ebenfalls den exakten Vergleich erschwert.

VALSECCHI et al. (2011) haben die Test-Retest-Reliabilität des *Temperament Test* untersucht. Zwölf Hunde konnten bereits am ersten Wesenstest aufgrund starker Angst nicht teilnehmen, da keine Interaktion möglich war; die Testung nach 60 Tagen ergab das gleiche

Problem, daher wurden diese Hunde aus der Studie exkludiert. Somit erfolgte für nur 44 *short term* Hunde (Hunde, welche erst kurzfristig im Tierheim untergebracht waren) die Evaluierung der Zuverlässigkeit des Tests bei erneuter Testung des Hundes in derselben Umgebung und erfolgte somit sowohl 20, als auch 60 Tage nach Aufnahme im Tierheim. Die Untersuchung ergab signifikante Korrelation zwischen den Gesamtpunktescores der beiden Tests. Die Test-Retest-Reliabilität innerhalb des Tierheimaufenthaltes war hoch, trotz möglicher Lerneffekte.

KIS et al. (2014) haben die Ergebnisse des Wesenstests *nach KLAUSZ et al.* von Hunden verglichen, wenn dieser ein bis zwei Tage oder zwei Wochen nach Aufnahme im Tierheim durchgeführt wurde. Fünfundneunzig Hunde wurden nach ein oder zwei Tagen im Tierheim getestet, davon konnten 25 Hunde nach zwei Wochen erneut getestet werden. Die Übereinstimmung für Aggression im Einzeltest „*Futter wegnehmen*“ ist schwächer als die der anderen Messparameter und aller Messparameter in den anderen Einzeltests. Würde man diesen Einzeltest exkludieren, wäre die Test-Retest-Reliabilität hoch.

MORNEMENT et al. (2014) haben 48 Hunde im *B.A.R.K.* Wesenstest getestet, welche seit mindestens drei Tagen im Tierheim waren. Alle Hunde wurden zu zwei Gelegenheiten getestet. Der erste Test wurde simultan durch zwei unabhängige BeobachterInnen beurteilt. Wiederholt wurde der Test nach 24 Stunden durch eine dieser beiden Personen in derselben Umgebung. Die höchste Test-Retest-Reliabilität erreichte der Messparameter *Angst. Nervosität* und *Aktivitätslevel* hatten moderate und statistisch signifikante Korrelationen zwischen den Tests. *Freundlichkeit* und *Fügsamkeit* schnitten am schlechtesten ab, aber dennoch gab es signifikante Korrelationen (siehe Tab. 30). Insgesamt war die Test-Retest-Reliabilität relativ schwach, obwohl der zeitliche Abstand zwischen den Tests nur 24 Stunden betrug.

FRANK (2007) hat 23 der 70 initialen Hunde dem Wiederholungstest nach vier Wochen unterzogen. Bei einem Hund (4.3%) konnte der Zwinger nicht betreten werden, es war einer der elf Hunde, die den Eingangstest abschlossen, aber auffällig waren. Bei einem Hund (4.3%) musste der Test abgebrochen werden, es war einer der acht Hunde, bei denen bereits der Eingangstest abgebrochen werden musste. Zwei der Hunde (9.5%) wurden im Wiederholungstest auffällig (Level 3). FRANK berichtet von einer positiven Verhaltensänderung, da im Gegensatz zu ca. 27% der Hunde im Eingangstest nur 17.4% der Hunde im

Wiederholungstest auffälliges Verhalten zeigten. Lediglich der Einzeltest „auf den Rücken drehen“ wies im Wiederholungstest einen Anstieg von aggressivem Verhalten gegenüber der Testerin auf. Die Test-Retest-Reliabilität wurde nicht im Sinne der Test-Retest-Reliabilität gemessen, sondern zum Zwecke die Verhaltensentwicklung zu erfassen.

Einige AutorInnen haben den Wiederholungstest im neuen Zuhause abgehalten. Die Sinnhaftigkeit ist fraglich, da es um die Wesensbeurteilung im Tierheim geht und ebendiese eine möglichst hohe Wiederholpräzision aufweisen sollte.

So wurden 39 der 236 Hunde im Wesenstest der RSPCA durch POULSEN (2010) und eine/n AssistentIn etwa 80 Tage nach der Vermittlung noch einmal getestet, direkt im neuen Zuhause. Die Situation wurde dem ersten Wesenstest bestmöglich nachgestellt, in Abwesenheit der/des NeubesitzerIn. Zwei der Testpunkte wurden im Wiederholungstest exkludiert (*Verhalten beim Alleinbleiben, Reaktion auf andere Hunde*). Die Ergebnisse der Wesenstests im Tierheim haben mit denen der Wiederholungstests nur in jenen Einzeltests korreliert, in denen kein direkter Kontakt mit den PrüferInnen bestand (*Futterverteidigung, Reaktion auf Bewegung/ Stimme/ Spielzeug, Spielversuch*). Die Ergebnisse des Futteraggressionstests haben in beiden Wesenstests stark korreliert.

VALSECCHI et al., (2011) haben für den *Temperament Test* zusätzlich zum Wiederholungstest im Tierheim die Zuverlässigkeit des Tests bei erneuter Testung des Hundes im neuen Zuhause evaluiert, vier Monate nach Vermittlung. Diese Testwiederholung wurde bei 34 Hunden durchgeführt, davon wurden 16 Hunde bereits zweimal im Tierheim getestet und 18 Hunde nur einmalig nach 20 Tagen im Tierheim. Der Test im neuen Zuhause erfolgte unter Anwesenheit der neuen Besitzerin/ des neuen Besitzers, es wurden aber nur 16 von 22 Einzeltests durchgeführt. Die Untersuchung ergab signifikante Korrelationen zwischen dem Test nach 20 Tagen und dem Test im neuen Zuhause. Die Studie sieht ihre Besonderheit darin, erstmals eine Testwiederholung im neuen Zuhause unter Anwesenheit der Neubesitzerin/ des Neubesitzers durchzuführen. Die Korrelation zwischen dem Wesenstest im Tierheim und dem im neuen Zuhause war signifikant geringer als zwischen den beiden Tests im Tierheim. Der höhere („bessere“) Punktescore im neuen Zuhause spricht für ein positiveres Verhalten nach der Vermittlung und würde wiederum für einen verminderten positiven prädiktiven Wert bei Problemverhalten im neuen Zuhause sprechen. Die Frage stellt sich also, ob es das Testergebnis ist, das vorhergesagt werden soll oder das

Alltagsverhalten im neuen Heim – ein allgemeines Dilemma bei der Vorhersagbarkeit von Wesenstests.

Auch KIS et al. (2014) haben Hunde im neuen Zuhause getestet (Wesenstest *nach KLAUSZ et al.*), aber lediglich um diese Ergebnisse mit jenen des zweiten Tests im Tierheim vergleichen zu können. Hierfür wurden analog zum zweiten Test im Tierheim 25 der 70 vermittelten Hunde ebenfalls nach zwei Wochen erneut getestet (zufällige Auswahl) und mit den Ergebnissen der 25 noch im Tierheim befindlichen Hunden verglichen. Es wurden keine signifikanten Unterschiede gefunden.

Für die Test-Retest-Reliabilität wurden je nach Studie völlig unterschiedliche statistische Messgrößen angewandt. Daher besteht eine eingeschränkte Vergleichbarkeit. Teilweise wurden innerhalb einer Studie bei der Messung unterschiedliche Tests verwendet (*gepaarter t-Test, Wilcoxon-Vorzeichen-Rang-Test* und *Chi-Quadrat-Test* in KIS et al., 2014). In einer Studie wurde der *Pearson-Korrelationskoeffizient* angewandt (MORNEMENT et al., 2014). In einer anderen Studie wurde der *Spearman'sche Rangkorrelationskoeffizient* verwendet (VALSECCHI et al., 2011) und in wieder einer anderen Studie wurde der *Kendall-Rangkorrelationskoeffizient* verwendet (POULSEN et al., 2010). In einer weiteren Studie wurde *Cohen's Kappa* verwendet (BENNETT et al., 2015).

Auch in den Wesenstests unterschieden einige AutorInnen zwischen *Angst* und *Furcht* (MORNEMENT et al., 2014).

Tab. 30: Test-Retest-Reliabilität der Wesenstests

Wesenstest	AutorInnen, die dieses Kriterium untersuchten	Abstand zwischen den Testungen	Test-Retest-Reliabilität
nach KLAUSZ et al.	KIS et al., 2014	2 Wochen	<p><i>Angst/Unterwerfung, Furcht/Unwohlsein</i> im Test „<i>Freundliche Begrüßung</i>“: keine Unterschiede</p> <p><i>Futterverteidigung</i>: Zunahme von <i>Aggression</i> und <i>physischer Kontakt</i> von Test zu Retest</p> <p><i>Angst/Unterwerfung</i> im Test „<i>bedrohliche Annäherung</i>“: keine Unterschiede</p>

			<i>Furcht/Unwohlsein im Test „bedrohliche Annäherung“</i> : Zunahme
B.A.R.K.	MORNEMENT et al., 2014	24 Stunden	<i>Angst</i> : hohe Übereinstimmung ( $r = 0.82$ ; $P < 0.01$ ) <sup>6</sup>  <i>Freundlichkeit</i> : deutlicher Zusammenhang ( $r = 0.53$ ; $P < 0.01$ )  <i>Aktivität</i> : deutlicher Zusammenhang ( $r = 0.69$ ; $P < 0.01$ )  <i>Gehorsam</i> : deutlicher Zusammenhang ( $r = 0.57$ ; $P < 0.01$ )  <i>Furcht</i> : deutlicher Zusammenhang ( $r = 0.71$ ; $P < 0.01$ )
Temperament test	VALSECCHI et al., 2011	Retest 1: 40 Tage  Retest 2: nach Adoption	Retest 1: deutlicher Zusammenhang ( $r_s = 0.58$ ; $P < 0.001$ ) <sup>7</sup>  Retest 2: mäßiger Zusammenhang ( $r_s = 0.44$ ; $P < 0.01$ )
RSPCA Wesenstest	POULSEN et al., 2010	80 Tage nach Adoption	Einzeltests mit direktem Kontakt mit PrüferInnen: keine signifikante Übereinstimmung  <i>Futterverteidigung</i> : moderate Übereinstimmung $\tau^8 = 0.32$ ( $P = 0.02$ )  <i>Reaktion auf Stimme und Bewegung</i> : moderate Übereinstimmung $\tau = 0.26$ ( $P = 0.03$ )  <i>Reaktion auf Spielzeug und Spielaufforderung</i> : starke Übereinstimmung $\tau = 0.49$ ( $P < 0.01$ )
SAFER™	BENNETT et	3 Tage	<i>Augenkontakt</i> :

6 *Korrelationskoeffizient nach Pearson*: 0.0-0.2 kein bis geringer Zusammenhang, 0.21-0.5 schwacher bis mäßiger Zusammenhang, 0.51-0.8 deutlicher Zusammenhang, 0.81-1.0 hoher bis perfekter Zusammenhang (<https://www.medistat.de/glossar/korrelation-assoziaton/korrelationskoeffizient-nach-pearson/>, 2021)

7 *Rangkorrelationskoeffizient nach Spearman*: 0.0-0.2 kein bis geringer Zusammenhang, 0.21-0.5 schwacher bis mäßiger Zusammenhang, 0.51-0.8 deutlicher Zusammenhang, 0.81-1.0 hoher bis perfekter Zusammenhang (<https://www.medistat.de/glossar/korrelation-assoziaton/rangkorrelationskoeffizient-nach-spearman/>, 2021)

8 *Kendall-Rangkorrelationskoeffizient*: 0-0.06 keine Übereinstimmung, 0.07-0.2 schwache Übereinstimmung, 0.21-0.34 moderate Übereinstimmung, ab 0.35 starke Übereinstimmung (<https://www.spss-tutorials.com/kendalls-tau/>)

	al., 2015		ausreichende Übereinstimmung $\kappa^9 = 0.28$ (95% CI <sup>10</sup> )  <i>Kneiftest:</i> ausreichende/hohe Übereinstimmung <sup>11</sup> $\kappa = 0.22/0.78$ (95% CI)  <i>Hundeverträglichkeit:</i> ausreichende Übereinstimmung $\kappa = 0.33$ (95% CI)  <i>Futterverteidigung:</i> moderate Übereinstimmung $\kappa = 0.50$ (95% CI)  <i>Berührungssensitivität:</i> moderate Übereinstimmung $\kappa = 0.59$ (95% CI)
nach FRANK	FRANK, 2007	4 Wochen	positive Verhaltensänderung (27% vs. 17.4% der Hunde auffällig)  Anstieg der Aggression bei „auf den Rücken drehen“

### 3.3.3.2.2 Inter-Rater-Reliabilität von Wesenstests

Um die Inter-Rater-Reliabilität zu evaluieren, ist es notwendig, dass mehrere BeurteilerInnen unabhängig voneinander denselben Wesenstest nach dem vorgegebenen Bewertungsschema beurteilen.

Die Inter-Rater-Reliabilität wurde untersucht für den Wesenstest nach *SAFER™* (WEISS, 2002), für den *Temperament Test* (VALSECCHI et al., 2011), für den B.A.R.K. Wesenstest (MORNEMENT et al., 2014), für den Wesenstest der *ASPCA®* (MATTERS, 2016), für den Wesenstest nach *KLAUSZ et al.* (KIS, 2014) und den Wesenstest nach *Dogs Trust* (DIESEL et al., 2008), siehe Tab. 26.

9 *Cohen's Kappa*: 0-0.2 schwache Übereinstimmung, 0.21-0.4 ausreichende Übereinstimmung 0.41-0.6 moderate Übereinstimmung, 0.61-0.8 erhebliche Übereinstimmung, 0.81-1.0 hohe Übereinstimmung (MATTERS, 2016)

10 CI = Konfidenzintervall

11 Die Hunde wurden zweimal gekniffen, beim zweiten Mal war die Übereinstimmung stärker

Obwohl sich in der Studie zum Wesenstest der *RSPCA* (POULSEN et al., 2010) um eine Angleichung der Beurteilungsweise zwischen den PrüferInnen und POULSEN (die den Wiederholungstest durchgeführt hat) bemüht wurde, gab es hier keine Untersuchung zur Inter-Rater-Reliabilität. In der Studie von CLAY et al. (2020a) zum Wesenstest der *RSPCA* wurde erwähnt, dass eine Untersuchung auf Inter-Rater-Reliabilität aufgrund von Personalwechseln nicht möglich war. Die Hunde in der Studie von LUCIDI et al. (2015) wurden anschließend an den *Ethotest* weiters von einer Gruppe von PrüferInnen der *Delta Society* getestet, die auf die Selektion von Hunden für Therapiehundeprogramme spezialisiert war. Die Inter-Rater-Reliabilität wurde als stark angegeben, da die Übereinstimmung zwischen dem *Ethotest* und der Evaluierung durch die *Delta Society* groß war. Allerdings wurde deren Evaluierungsverfahren nicht genau beschrieben und dürfte daher abweichen vom *Ethotest*. Daher kann man diese Referenz nicht als Inter-Rater-Reliabilität bezeichnen. In weiteren Studien wurde zwar durchaus die Notwendigkeit einer hohen Inter-Rater-Reliabilität zwar erwähnt, aber ebenfalls ohne explizite Untersuchung.

Für den Wesenstest nach *SAFER™* wurden die Videos aller Tests jeweils von drei Personen ausgewertet, um eine möglichst hohe Inter-Rater-Reliabilität zu erreichen (WEISS et GREENBERG, 1997). Die drei BeurteilerInnen haben sich letztendlich getroffen und Unklarheiten geklärt. Die tatsächliche Inter-Rater-Reliabilität wurde leider auch hier nicht gemessen. Es hieß, die externe Person, welche bei Uneinigkeiten zugezogen werden konnte, war nicht notwendig. Bei WEISS et al. (2002) hatten die TrainerInnen vor der Reevaluierung in Phase 4 einen zehnteiligen Fragebogen auszufüllen. Pro Hund gab es zwei TrainerInnen, die unabhängig voneinander die Fragen beantworteten. Diskrepanzen wurden diskutiert und gegebenenfalls gemittelt. Die Inter-Rater-Reliabilität wurde als sehr hoch angegeben, genaue Daten wurden allerdings nicht erhoben.

VALSECCHI et al. (2011) haben in Phase 1 ihrer Studie mit 19 *long-term* Hunden (Hunde, die bereits über ein Jahr im Tierheim lebten) die Inter-Rater-Reliabilität des *Temperament Test* gemessen. Hierfür wurde ein Vergleich der Auswertung des gleichen Videos durch zwei verschiedene TesterInnen (Testerin A weiblich, Tester B männlich) vorgenommen. Die Untersuchung ergab für *Verspieltheit* nahezu perfekte Übereinstimmung, für *Verhalten im Zwinger* und *kognitive Fähigkeiten* erhebliche Übereinstimmung, für *Umgänglichkeit mit Menschen*, *Umgänglichkeit mit Artgenossen* und *Reaktivität* moderate Übereinstimmung und



für *Fügsamkeit an der Leine* nur ausreichende Übereinstimmung (siehe Tab. 31). Für *Fügsamkeit an der Leine* gab es keine signifikante Korrelation zwischen Testerin A und Tester B. Die Inter-Rater-Reliabilität war grundsätzlich zufriedenstellend, allerdings gab es Testkategorien mit einem niedrigeren Level an Übereinstimmung, beispielsweise *Umgänglichkeit mit Menschen und Hunden*, welche zu den relevantesten Kategorien zählen dürften.

Die höchste Inter-Rater-Reliabilität des B.A.R.K. Wesenstests zwischen zwei unabhängigen BeurteilerInnen bei MORNEMENT et al. (2014) erreichte der Messparameter *Angst* in allen Einzeltests. Die niedrigste Inter-Rater-Reliabilität erreichte der Messparameter *Freundlichkeit*, aber die Ergebnisse korrelierten dennoch signifikant und stimmten mindestens moderat überein (siehe Tab. 31). Insgesamt war die Inter-Rater-Reliabilität moderat bis stark, aber fast alle Korrelationen waren statistisch signifikant. Nur vier Korrelationen waren statistisch nicht signifikant (beispielsweise der Messparameter *Nervosität* im Einzeltest „*Treffen eines freundlichen Hundes des anderen Geschlechts*“). Die prozentuale Übereinstimmung der beiden BeurteilerInnen reichte von 61% bis 100%, am besten schnitten *Freundlichkeit* (98%) und *Fügsamkeit* (94%) ab, am schlechtesten *Nervosität* mit 83%.

MATTERS (2016) untersuchte die Inter-Rater-Reliabilität für den Wesenstest der ASPCA®. Zwei der vier Beurteilerinnen wendeten den Wesenstest regelmäßig an (*current raters*) und die anderen beiden haben ihn zwar mitentwickelt, aber führten ihn zum Zeitpunkt der Untersuchung nicht selbst durch (*non-current raters*). Für die Inter-Rater-Reliabilität wurden alle Beurteilerinnen einzeln miteinander verglichen (1 + 2; 1 + 3; 1 + 4; 2 + 3; 2 + 4; 3 + 4 → ergibt sechs *Kombinationen*). Es wurden weiters die Übereinstimmungen nach den 29 Verhaltenskategorien aufgeschlüsselt. Die Übereinstimmungen im Allgemeinen waren gut bis exzellent. Die Inter-Rater-Reliabilität zwischen den beiden „*current*“ Beurteilerinnen war signifikant höher als bei allen anderen Kombinationen. In einigen der 29 Situationen waren die Übereinstimmungen zwischen den Beurteilerinnen nicht signifikant, beispielsweise bei der Beurteilung des Verhaltens allein in Außenzwinger oder in fremder Umgebung.

In der Studie von KIS et al. (2014) hat Prüferin 1 den Wesenstest nach KLAUSZ et al. durchgeführt und Prüferin 2 hat den Wesenstest gefilmt. Aufgrund dieser Tatsache konnte die Inter-Rater-Reliabilität untersucht werden und wurde in der Studie für alle Einzeltests als hoch berechnet.

DIESEL et al. (2008) haben die Inter-Rater-Reliabilität des Wesenstest nach *Dogs Trust* untersucht. Die Übereinstimmungen der Beurteilungen des Tierheimpersonals variierten je nach Messparameter und Testsituation zwischen schwach (beispielsweise für *angst-aggressives Verhalten beim Aufeinandertreffen mit einem anderen Artgenossen*) und moderat (beispielsweise für *nervöses Verhalten beim Annähern an den Zwinger*). Die Übereinstimmung zwischen jenen PrüferInnen war größer, welche offizielles entsprechendes Training genossen haben oder über acht Jahre Erfahrung mitbrachten, was bei 17 von 40 Personen der Fall war. Für den Messparameter „*gleichgültiges Verhalten*“ war die Reliabilität insgesamt sehr niedrig, was dafür spricht, dass dieser aus Gründen der Subjektivität und Schwierigkeit in seiner Erkennung eventuell ungeeignet ist. Selten beobachtete Verhaltensweisen, wie beispielsweise „*Angstaggression im allgemeinen Handling des Hundes*“, gehen ebenfalls mit einer schwachen Übereinstimmungen einher, siehe Tab. 31. Leider ist die Stichprobengröße in einigen Situationen sehr klein ausgefallen, so wurden beispielsweise nur sechs der 20 Hunde hinsichtlich ihres Verhaltens beim Treffen von Artgenossen beurteilt.

Die Inter-Rater-Reliabilität wird in einigen Studien in *Cohens Kappa* angegeben (KIS et al., 2014; VALSECCHI et al., 2011; MATTERS, 2016; DIESEL et al., 2008). In anderen Studien wird sie durch den *Pearson-Korrelationskoeffizienten* angegeben (MORNEMENT et al., 2014; MATTERS, 2016). Aufgrund dessen sollte beim direkten Vergleich der Messgrößen die mangelnde Einheitlichkeit beachtet werden.

Tab. 31: Inter-Rater-Reliabilität der Wesenstests

Wesenstest	AutorInnen, die dieses Kriterium untersuchten	BeurteilerInnen	Inter-Rater-Reliabilität
nach KLAUSZ et al.	KIS et al., 2014	2 Frauen (Prüferin, Filmerin)	hohe bis perfekte Übereinstimmung $\kappa^{12} = 1.0$ für <i>Angst/Unterwerfung, Aggression, Kontakt</i> $\kappa = 0.81$ für <i>Furcht/Unwohlsein</i>
B.A.R.K.	MORNEMENT et al., 2014	2 BeurteilerInnen	Angst/Furcht: hoher Zusammenhang ( $r = 0.95/0.92$ ; $P < 0.01$ ) <sup>13</sup> Gehorsam:

<sup>12</sup> *Cohen's Kappa*: 0-0.2 schwache Übereinstimmung, 0.21-0.4 ausreichende Übereinstimmung 0.41-0.6 moderate Übereinstimmung, 0.61-0.8 erhebliche Übereinstimmung, 0.81-1.0 hohe Übereinstimmung (MATTERS, 2016)

			<p>hoher Zusammenhang (<math>r = 0.87</math>; <math>P &lt; 0.01</math>)</p> <p>Aktivitätslevel: deutlicher Zusammenhang (<math>r = 0.72</math>; <math>P &lt; 0.01</math>)</p> <p>Freundlichkeit: deutlicher Zusammenhang (<math>r = 0.70</math>; <math>P &lt; 0.01</math>)</p>
Temperament test	VALSECCHI et al., 2014	<p>Testerin A weiblich</p> <p>Tester B männlich</p>	<p>Verspieltheit: hohe Übereinstimmung <math>\kappa = 0.90</math> (<math>P &lt; 0.001</math>)</p> <p>Verhalten im Zwinger: hohe Übereinstimmung <math>\kappa = 0.72</math> (<math>P &lt; 0.001</math>)</p> <p>Umgänglichkeit: moderate Übereinstimmung; Artgenossen <math>\kappa = 0.42</math> (<math>P &lt; 0.01</math>) Menschen <math>\kappa = 0.44</math> (<math>P &lt; 0.01</math>)</p> <p>Reaktivität: moderate Übereinstimmung <math>\kappa = 0.57</math> (<math>P &lt; 0.001</math>)</p> <p>Fügsamkeit an der Leine: keine signifikanten Übereinstimmungen</p> <p>Kognitive Fähigkeiten: hohe Übereinstimmung <math>\kappa = 0.80</math> (<math>P &lt; 0.001</math>)</p>
ASPCA®	MATTERS, 2016	<p>2 <i>current</i> Beurteilerinnen</p> <p>2 <i>non-current</i> Beurteilerinnen</p>	<p>Zwischen den <i>current</i> Beurteilerinnen: hoher Zusammenhang <math>r = 0.87/0.83^{14}</math> (<math>P &lt; 0.001/0.01</math>) <math>\kappa = 0.85/0.81</math> (<math>P &lt; 0.001</math>)</p> <p>in allen anderen Kombinationen von BeurteilerInnen: erhebliche Übereinstimmung <math>r = 0.67</math> bis <math>0.77</math> (<math>P &lt; 0.001/0.01</math>) <math>\kappa = 0.63</math> bis <math>0.72</math> (<math>P &lt; 0.001</math>)</p> <p>erhebliche bis fast perfekte Übereinstimmung in 22 von 29 Testsituationen</p>

13 *Korrelationskoeffizient nach Pearson*: 0.0-0.2 kein bis geringer Zusammenhang, 0.21-0.5 schwacher bis mäßiger Zusammenhang, 0.51-0.8 deutlicher Zusammenhang, 0.81-1.0 hoher bis perfekter Zusammenhang (<https://www.medistat.de/glossar/korrelation-assoziati/korrelationskoeffizient-nach-pearson/>, 2021)

14 Der erste Wert bezieht sich auf die erste Evaluierung, der zweite Wert bezieht sich auf die zweite Evaluierung

			ausreichende bis moderate Übereinstimmung in 7 von 29 Testsituationen
SAFER™	WEISS et al., 2002	2 TrainerInnen	„sehr hoch“, keine detaillierten Daten verfügbar
Dog´s Trust	DIESEL et al., 2008	40 BeurteilerInnen, davon 17 sehr erfahren	schwache bis maximal moderate Übereinstimmung, nur innerhalb der Gruppe der sehr erfahrenen BeurteilerInnen erhebliche Übereinstimmung in 3 von 18 Verhaltensweisen <sup>15</sup>  Annäherung an den Zwinger:  BeurteilerInnen mit weniger als 8 Jahren Erfahrung: $\kappa = 0.04$ ( <i>Gleichgültigkeit</i> ) bis $0.58$ ( <i>Nervosität</i> ) BeurteilerInnen mit über 8 Jahren Erfahrung: $\kappa = 0.12$ ( <i>Gleichgültigkeit</i> ) bis $0.72$ ( <i>Nervosität</i> )  Allgemeines Handling und Pflegen:  BeurteilerInnen mit weniger als 8 Jahren Erfahrung: $\kappa = 0.03$ ( <i>Aggression</i> ) bis $0.46$ ( <i>Nervosität</i> ) BeurteilerInnen mit über 8 Jahren Erfahrung: $\kappa = 0.06$ ( <i>Aggression</i> ) bis $0.54$ ( <i>Nervosität</i> )  Reaktion auf einen Artgenossen:  BeurteilerInnen mit weniger als 8 Jahren Erfahrung: $\kappa = 0.11$ ( <i>Aggression</i> ) bis $0.37$ ( <i>Aufregung</i> ) BeurteilerInnen mit über 8 Jahren Erfahrung: $\kappa = 0.15$ ( <i>Aggression</i> ) bis $0.47$ ( <i>Aufregung</i> )

### 3.3.3.2.3 Intra-Rater-Reliabilität von Wesenstests

Um die Intra-Rater-Reliabilität eines Wesenstests zu beurteilen, ist es notwendig, dass die/der selbe BeurteilerIn denselben Wesenstest zu verschiedenen Zeitpunkten beurteilt.

Die Intra-Rater-Reliabilität wurde lediglich für den *Temperament Test* (VALSECCHI et al., 2011), den Wesenstest der ASPCA® (MATTERS, 2016) und den Wesenstest nach *Dogs*

<sup>15</sup> Beurteilt wurde in drei Testsituationen nach jeweils 6 möglichen Verhaltensweisen: *angst aggressives Verhalten*, *nervöses Verhalten*, *gleichgültiges Verhalten*, *erwünschtes Verhalten*, *aufgeregtes Verhalten*, *offensiv-aggressives Verhalten*.

*Trust* (DIESEL et al., 2008) untersucht. Zu allen weiteren Wesenstests fehlen dahingehend Informationen.

VALSECCHI et al. (2011) haben in ihrer Studie zum *Temperament Test* die Intra-Rater-Reliabilität mit 19 *long-term* Hunden gemessen. Hierfür wurden zwei Auswertungen der gleichen Testerin/ des gleichen Testers (Testerin A nach einem Monat, Tester B nach 16 Monaten) verglichen. Für Testerin A ergab die Untersuchung in fast allen Testkategorien erhebliche bis perfekte, außer für „*Fügsamkeit an der Leine*“ nur moderate Übereinstimmung (siehe Tab. 32). Für Tester B ergab die Untersuchung in allen Testkategorien moderate bis nahezu perfekte Übereinstimmung (Übereinstimmung insgesamt etwas geringer als bei Testerin A). Die Intra-Rater-Reliabilität war grundsätzlich zufriedenstellend, allerdings gab es Testkategorien mit einem niedrigeren Level an Übereinstimmung, beispielsweise „*Umgänglichkeit mit Menschen und Hunden*“, welche zu den relevantesten Kategorien zählen dürften.

In der Studie von MATTERS (2016) zum Wesenstest der ASPCA® sah sich jede/r BeurteilerIn („*current*“ und „*non-current*“) die Videoaufnahmen eines Wesenstests von zwölf Hunden zweimal im Abstand von vier bis fünf Wochen an. Der Ton wurde abgeschaltet, damit die Kommentare der testenden Personen nicht zu hören waren. Für die Intra-Rater-Reliabilität wurden die beiden Beurteilungen für jede Beurteilerin für alle 29 Situationen verglichen. Die Übereinstimmungen waren im Allgemeinen gut bis exzellent. Überraschenderweise hatte eine der „*non-current*“ Beurteilerinnen die höchste Intra-Rater-Reliabilität.

DIESEL et al. (2008) haben die Intra-Rater-Reliabilität des Wesenstest nach *Dogs Trust* untersucht. Die Beurteilung wurde nach einer Woche bis sechs Monaten mit demselben Video vom Wesenstest wiederholt, 18 Personen haben daran teilgenommen. Die Intra-Rater-Reliabilität war generell moderat bis hoch in allen Fällen, nur bezüglich einiger Verhaltensweisen beim Treffen von Artgenossen gab es nur schwache Übereinstimmungen. Die Intra-Rater-Reliabilität ist jedenfalls stärker als die Inter-Rater-Reliabilität, allerdings war auch die Anzahl der TeilnehmerInnen bei dieser Untersuchung geringer (18 vs. 40).

Für die Darstellung der Intra-Rater-Reliabilität wurde einheitlich *Cohen's Kappa* verwendet (VALSECCHI et al., 2011; MATTERS, 2016; DIESEL et al., 2008). Andere AutorInnen haben zusätzlich den *Pearson-Korrelationskoeffizienten* verwendet (MATTERS, 2016). Die

Vergleichbarkeit ist in diesem Fall durchaus gegeben, da alle AutorInnen *Cohen's Kappa* gemessen haben, siehe Tab. 32.

Tab. 32: Intra-Rater-Reliabilität der Wesenstests

Wesenstest	AutorInnen, die dieses Kriterium untersuchten	Abstand zwischen den Beurteilungen	Intra-Rater-Reliabilität
Temperament test	VALSECCHI et al., 2011	Testerin A: 1 Monat Tester B: 16 Monate	<p>Testerin A: erhebliche bis perfekte Übereinstimmung in 6 von 7 Testkategorien (1 von 7 moderate Übereinstimmung)</p> <p>Tester B: moderate bis nahezu perfekte Übereinstimmung in allen Testkategorien</p> <p><i>Verhalten im Zwinger</i>: <math>\kappa^{16} = 1.00/0.80^{17}</math> (<math>P &lt; 0.001</math>)</p> <p><i>Umgänglichkeit mit Hunden</i>: <math>\kappa = 0.69/0.58</math> (<math>P &lt; 0.001</math>)</p> <p><i>Umgänglichkeit mit Menschen</i>: <math>\kappa = 0.76/0.55</math> (<math>P &lt; 0.001</math>)</p> <p><i>Verspieltheit</i>: <math>\kappa = 0.96/0.97</math> (<math>P &lt; 0.001</math>)</p> <p><i>Reaktivität</i>: <math>\kappa = 0.83/0.86</math> (<math>P &lt; 0.001</math>)</p> <p><i>Fügsamkeit an der Leine</i>: <math>\kappa = 0.46/0.79</math> (<math>P &lt; 0.01 / &lt; 0.001</math>)</p> <p><i>Kognitive Fähigkeiten</i>: <math>\kappa = 0.92/0.78</math> (<math>P &lt; 0.001</math>)</p>
ASPCA®	MATTERS, 2016	4 – 5 Wochen (4 BeurteilerInnen)	<p>erhebliche bis fast perfekte Übereinstimmung</p> <p><math>r = 0.86/0.88</math> (<math>P &lt; 0.001</math>)<sup>18</sup> <math>\kappa = 0.83/0.86</math> (<math>P &lt; 0.001</math>)<sup>19</sup></p> <p><math>r = 0.79/0.97</math> (<math>P &lt; 0.001</math>) <math>\kappa = 0.71/0.96</math> (<math>P &lt; 0.001</math>)<sup>20</sup></p>

16 *Cohen's Kappa*: 0-0.2 schwache Übereinstimmung, 0.21-0.4 ausreichende Übereinstimmung 0.41-0.6 moderate Übereinstimmung, 0.61-0.8 erhebliche Übereinstimmung, 0.81-1.0 hohe Übereinstimmung (MATTERS, 2016)

17 Der erste Wert bezieht sich auf Testerin A, der zweite Wert auf Tester B

18 *Korrelationskoeffizient nach Pearson*: 0.0-0.2 kein bis geringer Zusammenhang, 0.21-0.5 schwacher bis mäßiger Zusammenhang, 0.51-0.8 deutlicher Zusammenhang, 0.81-1.0 hoher bis perfekter Zusammenhang (<https://www.medistat.de/glossar/korrelation-assoziati/korrelationskoeffizient-nach-pearson/>, 2021)

19 Current raters

20 Non-current raters

Dog's Trust	DIESEL et al., 2008	1 Woche bis 6 Monate (18 BeurteilerInnen)	<p><i>Annäherung an den Zwinger:</i> ausreichende bis erhebliche Übereinstimmung <math>\kappa = 0.37</math> (<i>Gleichgültigkeit</i>) bis <math>0.79</math> (<i>Aggression</i>)</p> <p><i>Allgemeines Handling und Pflegen:</i> ausreichende bis hohe Übereinstimmung <math>\kappa = 0.35</math> (<i>Angst aggression</i>) bis <math>0.86</math> (<i>Aggression</i>)</p> <p><i>Reaktion auf einen Artgenossen:</i> schwache bis moderate Übereinstimmung <math>\kappa = 0.11</math> (<i>erwünschtes Verhalten</i>) bis <math>0.43</math> (<i>Aufregung</i>)</p>
-------------	------------------------	---	---

#### 3.3.3.2.4. Interne Konsistenz

Eine Untersuchung der internen Konsistenz erfolgte nur für den *Temperament Test* (VALSECCHI et al., 2011; BARNARD et al., 2019) und für den *Match-Up* Wesenstest (DOWLING-GUYER et al., 2011). Ein Vergleich ist bei dieser Messung möglich, da die interne Konsistenz einheitlich mit *Cronbach's Alpha* ausgedrückt wird.

VALSECCHI et al. (2011) haben im *Temperament Test* die Korrelation der Einzeltests mit den Messparametern gemessen für die beiden Hauptfaktoren *Umgänglichkeit mit Menschen* ( $\alpha = 0.75$ ) und *Umgänglichkeit mit Hunden* ( $\alpha = 0.82$ ) getestet, was eine hohe Konsistenz zwischen den Einzeltests ergab. BARNARD et al. (2019) haben die interne Konsistenz für die drei Hauptfaktoren *Umgänglichkeit mit Menschen* ( $\alpha = 0.83$ ), *Umgänglichkeit mit Hunden* ( $\alpha = 0.93$ ), sowie *Reaktivität* ( $\alpha = 0.65$ ) getestet. Sie ergab eine hohe Konsistenz zwischen den Einzeltests, nur für *Reaktivität* lag sie knapp unter dem akzeptablen Wert von 0.70. In einem der beiden teilnehmenden Tierheime war der Faktor *Umgänglichkeit mit Hunden* signifikant höher, allerdings wurden in diesem Tierheim die Hunde nach Möglichkeit paarweise gehalten.

Der Inhalt der Studie von DOWLING-GUYER et al. (2011) zum *Match-Up* Wesenstest war die Untersuchung der internen Konsistenz für die vier Messparameter. Achtunddreißig Verhaltensweisen (beispielsweise *Hochspringen*, *Piloerektion*, *Schwanzstellung hoch*,...) werden in jedem Einzeltest ausgewertet in *beobachtet/ nicht beobachtet*. Zwölf der 38 Verhaltensweisen wurden exkludiert, da sie nicht konsistent genug waren für die gewählten

Messparameter. Die interne Konsistenz wurde dann für die verbleibenden 26 Verhaltensweisen gemessen. Für alle Messparameter war die interne Konsistenz akzeptabel (*Angst*:  $\alpha = 0.733$ ; *Freundlichkeit/Verspieltheit*:  $\alpha = 0.700$ ; *Interesse/Erwartung*:  $\alpha = 0.746$ ), nur für *Aggression* lag sie knapp unter dem Mindestmaß ( $\alpha = 0.676$ ).

### **3.3.3.3 Validität von Wesenstests**

#### **3.3.3.3.1 Inhaltsvalidität**

Leider haben die meisten AutorInnen die Inhaltsvalidität nicht gemessen. Lediglich für den *Temperament Test* (VALSECCHI et al., 2011; BARNARD et al., 2019) wurde dies in die Studien einbezogen.

Bei VALSECCHI et al. (2011) wurde für 121 Hunde über Faktorenanalyse evaluiert, ob der Wesenstest eine gute Maßnahme zur Messung von Temperament ist. Aus den 22 Einzeltests wurden sechs Faktoren extrahiert: *Umgänglichkeit mit Menschen und Hunden*, *Verspieltheit*, *Fügsamkeit an der Leine*, *Stereotypien* und *kognitive Fähigkeiten*. Die Varianz war bei *Umgänglichkeit mit Menschen und Hunden* am größten. Vier der 22 Einzeltests konnten nicht signifikant in einem der sechs Faktoren zugeordnet werden. Auch BARNARD et al. (2019) haben die Inhaltsvalidität des *Temperament Tests* gemessen. Aus den 22 Einzeltests wurden fünf Faktoren extrahiert: *Umgänglichkeit mit Menschen und Hunden*, *Reaktivität*, *Anpassung an den Zwinger* und *Verspieltheit*. Die Varianz war ebenfalls bei *Umgänglichkeit mit Menschen und Hunden* am größten. Drei der 22 Einzeltests konnten keinem der sechs Faktoren signifikant zugeordnet werden.

#### **3.3.3.3.2 Kriteriumsvalidität/ prognostische Validität**

Als externes Kriterium wurde grundsätzlich das Verhalten im neuen Zuhause verwendet. VALSECCHI et al. (2011) haben allerdings in ihrer Studie zum *Temperament Test* ein anderes externes Kriterium hinzugezogen: drei Hundegruppen (10 x 3 *long-term* Hunde) mit unterschiedlichem Verhaltensprofil (*aggressiv*, *ängstlich*, *gesellig*) wurden verglichen. Diese Verhaltensprofile wurden durch TierheimtierärztInnen ausgestellt basierend auf Informationen zur Vorgeschichte und Verhaltensanamnese (*externes Kriterium*). Dabei bildete Gruppe 1 Hunde mit Bisszeugnis und Anamnese von offener Aggression gegenüber Menschen, Gruppe 2 Hunde mit Anamnese von Angst gegenüber Menschen und Gruppe 3



Hunde, die von TierärztInnen und Tierheimpersonal als freundlich und sozial eingeschätzt wurden. Alle Gruppen durchliefen den Wesenstest und wurden hinsichtlich der sieben Testkategorien verglichen. Es gab signifikante Unterschiede zwischen den Gruppen in den Testkategorien *Umgänglichkeit mit Menschen*, *Fügsamkeit an der Leine*, *Umgänglichkeit mit Artgenossen*, *Reaktivität* und *Verspieltheit*. Die Hunde der Gruppe 3 hatten signifikant höhere (bessere) Punktescores in allen sieben Testkategorien, verglichen mit den Hunden der anderen beiden Gruppen. Die Hunde der Gruppe 1 hatten im Vergleich zu den Hunden der Gruppe 2 signifikant niedrigere Punktescores in den Kategorien *Umgänglichkeit mit Menschen*, *Umgänglichkeit mit Artgenossen* und *Reaktivität*. Als Referenzwert wurde die Vorklassifikation in die drei Gruppen herangezogen und auf diesem Wege die Sensitivität des Wesenstests dafür abgeschätzt, diese drei Verhaltensweisen zu erkennen. Deutliches Potenzial ist in der Testkategorie *Umgänglichkeit mit Menschen* erkennbar, genaue Angaben zu Sensitivität, Spezifität, PPW und NPW lassen sich allerdings der Studie nicht entnehmen.

In weiterer Folge wird die prognostische Validität für das Verhalten im neuen Zuhause betrachtet. Diese wurde für die Wesenstests nach *VAN DER BORG et al. (1991)*, *Assess-A-Pet* (*CHRISTENSEN et al., 2007*; *BOLLEN et HOROWITZ, 2008*; *MCGUIRE et al., 2020*), nach *SAFER™* (*MOHAN-GIBBONS et al., 2012*), der *RSPCA* (*POULSEN et al., 2010*), *B.A.R.K.* (*MORNEMENT et al., 2014*; *MORNEMENT et al., 2015*), *Match-Up* (*MARDER et al., 2013*) und für den *Ethotest* (*STICCO et al., 2011*) untersucht, siehe Tab. 26. Zum *Temperament Test*, zum Wesenstest der *ASPCA®*, zum Wesenstest nach *KLAUSZ et al.* und nach *Dogs Trust* fehlen jegliche Untersuchungen zur prognostischen Validität.

In den ersten Studien zum Wesenstest nach *SAFER™* (*WEISS et GREENBERG, 1997*; *WEISS, 2002*) wurde nicht das Verhalten im neuen Zuhause, sondern in einem erneuten Wesenstest nach der Ausbildung zum Assistenzhund beurteilt. Der Studie von *MOHAN-GIBBONS et al. (2018)* zum *SAFER™*-Test können wir ebenfalls keine Informationen zur Validität entnehmen, da der Referenzwert nach Adoption auf selbstständigen Berichten durch die NeubesitzerInnen besteht und nicht auf einer Follow-Up-Studie. Die im *SAFER™* Wesenstest als „Futterverteidiger“ identifizierten Hunde in der Studie von *MOHAN-GIBBONS et al. (2012)* wurden mit entsprechendem Trainingsprogramm vermittelt, welches bereits während des Tierheimaufenthalts umgesetzt wurde. Da nur positiv-getestete Hunde Objekt der Studie waren und keine Informationen über die negativ-getesteten Hunde vorliegen, lässt

sich lediglich der positive prädiktive Wert berechnen. Dieser beträgt für Futtermverteidigung 10% (bei sechs Hunden wurden innerhalb der ersten drei Monate Verteidigen von Futter berichtet, 60 NeubesitzerInnen wurden mindestens einmal erreicht). Aber auch dieser Wert ist nicht ganz aussagekräftig, da nur Hunde involviert wurden, welche ausschließlich im spezifischen Einzeltest Futtermverteidigung gezeigt haben. Die Zahl der Hunde, die Futtermverteidigung und auch Aggression in anderen Einzeltests gezeigt haben, ist unbekannt. Allerdings ist die Fragestellung der Studie auf genau jene Hunde bezogen, die abgesehen vom Einzeltest auf Futtermverteidigung unauffällig sind.

In den Wesenstests nach *ZILOCCHI et al.* (2015) und nach *FRANK* (2007) wurde ebenfalls nicht das Verhalten im neuen Zuhause beurteilt, sondern ob der Hund vermittelt wurde oder nicht.

Die Befragung der NeubesitzerInnen durch *VAN DER BORG et al.* (1991) erfolgte zwischen ein und zwei Monaten nach Vergabe via Telefon, unter Verwendung der gleichen Fragebögen, mit denen das Tierheimpersonal die Hunde bereits beurteilt hat. Zusätzlich wurden die NeubesitzerInnen gezielt nach aufgetretenen Problemen mit den Hunden befragt. Durch diese drei Varianten der Beurteilung sollte die Validität der Wesenstests gemessen werden, (Problem-)verhalten vorauszusagen. Die Wahrscheinlichkeit des Auftauchens der Messparameter in den Einzeltests variierte sehr stark je nach Einzeltest. Die Befragungen der NeubesitzerInnen zeigten 17 Arten von Problemverhalten, wobei die Arten von Aggression für die Befragten schlecht differenzierbar waren. Die Wesenstests konnten etwa 75% der später aufgetretenen Problemverhalten korrekt voraussagen. Der Vorhersagewert der Tests bezüglich *Trennungsangst*, *Problemen beim Autofahren* und der *Verunreinigung des Hauses* war nicht sehr exakt. So konnte zum Beispiel *Verunreinigung des Hauses* in 33.3% der Fälle nur durch das Tierheimpersonal vorausgesagt werden und in 50.0% der Fälle gar nicht (weder durch den Wesenstest, noch durch das Personal). Auch *Probleme beim Autofahren* konnten in 38.5% der Fälle gar nicht vorausgesagt werden, waren also falsch-negativ in beiden Beurteilungsverfahren. Die Validität ergibt sich aus der korrekten Vorhersage über die Anwesenheit und Abwesenheit eines potentiellen Problemverhaltens in der Zukunft. *VAN DER BORG et al.* sahen daher die Priorität in der Berechnung der Sensitivität und des negativen prädiktiven Wertes. Die sechs häufigsten Problemverhalten wurden dahingehend untersucht (*Ziehen an der Leine*, *Aggression gegenüber Erwachsenen*,

*Ungehorsam, Trennungsangst, Aggression gegenüber Hunden und Probleme beim Autofahren*), siehe Tab. 33.

Für die Untersuchung der prognostischen Validität des *Assess-A-Pet* Wesenstests erfolgte das Telefoninterview durch CHRISTENSEN et al. (2007) durch eine Person, die vorher weder mit dem Hund, noch mit den AdoptantInnen zu tun hatte. Die Messparameter im Interview waren *Anspringen, Unsauberkeit, Trennungsangst, Bellen* und *Aggression*. Hierfür wurde eine Liste verwendet, die auf der Verhaltensklinik der *Cornell University* routinemäßig für Konsultationen angewandt wird und Fragen zu physischer Manipulation des Hundes, diversen Alltagssituationen und territorialem Verhalten enthält. Jeder Hund wurde anhand der Antworten der BesitzerInnen in eine von vier Gruppen klassifiziert: *keine Aggression/geringgradige Aggression* (nur Bellen), *mittelgradige Aggression* (Knurren oder Attackieren mit/ohne Bellen), *hochgradige Aggression* (Beißen/ Schnappen mit/ohne Knurren, Attackieren, Bellen). Knapp 41% der im Tierheim negativ getesteten Hunde zeigten dennoch Attackieren, Knurren, Schnappen, und/oder Beißen nach der Adoption (35% mgr. Aggression und 6% hgr. Aggression). Bezieht man „*geringgradige Aggression*“ mit ein, waren es sogar über 71%. Welches Bellen als Teil von Aggressionsverhalten einhergeht und welches Bellen nicht, ist allerdings kontrovers, da das Einschätzungsvermögen der NeubesitzerInnen diesbezüglich fraglich ist und die Bedeutung von Bellen stark von der Situation abhängt. Manche Typen von Aggression (territoriale Aggression, Beuteaggression und innerartliche Aggression) können demnach nicht zuverlässig im Test identifiziert werden. So zeigten 51.5% der vermittelten Hunde im neuen Zuhause territoriale Aggression, 16.7% innerartliche Aggression und 13.6% Beuteaggression. Hochgradige Aggression wurde nur in den Kategorien „territoriale Aggression“ und „Beuteaggression“ gezeigt. Sechs der Hunde (jeweils drei) zeigten *Aggression gegenüber dem/der BesitzerIn* oder *Ressourcenverteidigung gegenüber Menschen*, jeweils mit mittelgradigem Aggressionslevel. Da die positiv getesteten Hunde euthanasiert wurden, ist unklar, wie sie beispielsweise beim Streicheln, Anstarren oder Umarmen durch eine Bezugsperson anstatt einer fremden Testperson reagiert hätten und ob sie dabei auch Aggression gezeigt hätten. Aufgrund der Euthanasie aller Positiv-Getesteten können die Sensitivität, die Spezifität und der positive prädiktive Wert nicht gemessen werden, da der Referenzwert lediglich für negativ getestete Hunde vorhanden ist.

Auch BOLLEN et HOROWITZ (2008) untersuchten die prognostische Validität des *Assess-A-Pet* Wesenstests. Sechs Monate nach Adoption wurden Follow-Up-Anrufe getätigt, in denen Informationen über das Verhalten in Situationen ähnlich der Wesenstests erbeten wurden. Auch hier wurden drei Kategorien eingeteilt:

- „*nicht zurückgegeben/ zurückgegeben aus anderen Gründen als Aggression*“ (Referenzkategorie „*keine Probleme*“)
- „*zurückgegeben aufgrund von Verhaltensproblemen außer Aggression*“
- „*zurückgegeben aufgrund von Aggression*“.

Es wurde ein System entwickelt, das die Hunde in „*aggressiv*“ und „*nicht aggressiv*“ klassifiziert, basierend auf ihren demographischen Faktoren und den Ergebnissen des Wesenstests. Die Ergebnisse wurden retrospektiv mit den Verhaltenshistorien der Hunde verglichen, welche bei Aufnahme im Tierheim erhalten wurden. Laut der Studie war eine Bewertung der Sensitivität, Spezifität und Exaktheit der Klassifizierungstests möglich. Unter Sensitivität wurden die „*aggressiven*“ Abgabehunde verstanden, welche als aggressiv erkannt wurden im Wesenstest und unter Spezifität die „*nicht-aggressiven*“ Abgabehunde, welche im Wesenstest als nicht-aggressiv erkannt wurden. Den Referenzwert bildete die Vorgeschichte bei Abgabe. Da manche Testsysteme eine bessere Sensitivität und manche eine bessere Spezifität hatten, wurde ein dritter Wert, die Präzision (*accuracy*)  $P_C$  eingeführt, welcher die Wahrscheinlichkeit einer korrekten Klassifikation messen sollte, welcher sich folgendermaßen errechnet:  $P_C = \text{Sensitivität} \times p + \text{Spezifität} \times (1 - p^{21})$ . Die höchste Sensitivität (0.53) und Spezifität (0.89) wurde erreicht, wenn demographische Werte, Wesenstestergebnisse und der Interaktionseffekt einbezogen wurden. Bezieht man nur die demographischen Werte, bzw. das Alter und die Wesenstestergebnisse mit ein, erreicht man eine Sensitivität von 0.46 (bzw. 0.49) und eine Spezifität von 0.89 (bzw. 0.87). Die Präzision bewegt sich jeweils um 0.84. Die Sensitivität wurde dennoch als angemessen bezeichnet. Wird nur der Wesenstest betrachtet, ist die Sensitivität bei 0.90, da unter den Hunden mit Aggressionsvorgeschichte insgesamt 90% entweder *durchfallen*, *grenzwertig* oder *unsozial* waren und somit Aggression in irgendeiner Form zeigten. Wird nur der Wesenstest betrachtet, ist die Spezifität nur bei 0.27, da unter den Hunden ohne Aggressionsvorgeschichte nur 27% keinerlei Aggression im Wesenstest zeigten. Hunde, die den

---

21 „p“ beschreibt dabei die Prävalenz von Aggression in der Population der abgegebenen Hunde

Wesenstest bestanden haben, aber als *unsozial* klassifiziert wurden, waren assoziiert mit nicht-aggressiven Verhaltensproblemen beim Follow-Up. Der positive prädiktive Wert lässt sich hier nicht eruieren, da die meisten positiv getesteten Hunde euthanasiert wurden. Der negative prädiktive Wert in dieser Studie lag bei 0.92.

Einhundertneununddreißig Hunde wurden für die Untersuchung der prognostischen Validität des *Assess-A-Pet* Wesenstests durch MCGUIRE et al. (2020) im neuen Zuhause über Befragungen nach ein bis drei Monaten reevaluiert, davon wurden 20 Hunde (14%) im Tierheim als Ressourcenverteidiger identifiziert. Die NeubesitzerInnen wurden im Zuge der Adoption über ein eventuelles ressourcenverteidigendes Verhalten informiert und beraten. Die NeubesitzerInnen sollten die Verhaltensreaktionen ihres Hundes einordnen in „*keine sichtbaren Zeichen von Aggression*“/ „*Knurren*“/ „*Schnappen oder Beißen*“. Dies erfolgte für fünf verschiedene Situationen:

- beim Wegnehmen von Spielzeugen/ Knochen/ anderen hochgeschätzten Objekten
- beim Wegnehmen von Futter
- beim Zurückverlangen von Gegenständen oder Futter, was vom Hund geklaut wurde
- beim Annähern an den Hund, während er frisst
- beim Annähern an den Hund, während er auf seinem Lieblingsschlafplatz ist.

Die Befragung erfolgte durch *C-BARQ*, teilweise modifiziert nach MARDER et al. (2013). Wenn die Fragen bejaht wurden, wurde weiters danach gefragt, ob das Problem für die/den NeubesitzerIn gravierend ist und ob sie Schwierigkeiten damit haben, besitzergreifendes Verhalten zu vermeiden und zu managen. Für drei Situationen wurde ein signifikanter Zusammenhang zwischen dem Wesenstest im Tierheim und dem Berichten von Ressourcenverteidigung im neuen Zuhause gefunden:

- beim Wegnehmen von Spielzeugen, Knochen oder anderen hochgeschätzten Objekten → PPW 47.7%, NPW 88.9%,
- beim Wegnehmen von Futter → PPW 33.3% und NPW 97.4%
- beim Zurückverlangen von Gegenständen oder Futter, was vom Hund geklaut wurde → PPW 23.5% und NPW 95.5%.

Für die anderen beiden Situationen war der positive prädiktive Wert jeweils kleiner oder gleich 10% und der negative prädiktive Wert jeweils über 97%. Wenn man aus den Daten die Sensitivität berechnet, liegen die Werte je nach Situation zwischen 25% (*Annähern an den Lieblingsschlafplatz*) und 66.7% (*Wegnehmen von Futter*), für die Spezifität liegen die Werte zwischen 85.8% (*Annähern an den Lieblingsschlafplatz*) und 91.2% (*Wegnehmen von Spielzeug, Knochen oder anderen hochgeschätzten Objekten*). MCGUIRE et al. haben weiters alle drei Informationsquellen (VorbesitzerInnen, *Assess-A-Pet* Wesenstest und NeubesitzerInnen) zusammengefügt, allerdings gab es alle Informationen nur bei 44 Hunden. Bei 65.9% der Hunde stimmten alle drei Informationsquellen überein, was in der Studie als ausreichend bezeichnet wurde. Der Grad der Übereinstimmung war zwischen Wesenstest und NeubesitzerInnen leicht höher als zwischen VorbesitzerInnen und NeubesitzerInnen, bzw. VorbesitzerInnen und Wesenstest. Wenn man aus den Daten der Informationen der abgebenden BesitzerInnen die Sensitivität für das Auftreten von Ressourcenverteidigung berechnet liegt der Wert bei 71.4%, die Spezifität liegt bei 78.4%.

Für die Untersuchung der prognostischen Validität des Wesenstests der *RSPCA* wurde eine Befragung der NeubesitzerInnen durch POULSEN et al. (2010) zur Zufriedenheit mit dem adoptierten Hunden und zu etwaigen Verhaltensproblemen durchgeführt. Auch die im Wiederholungstest exkludierten Testpunkte wurden auf diesem Weg erfragt. Fünfundvierzig NeubesitzerInnen haben an der Befragung teilgenommen. Trotz einiger Verhaltensprobleme waren die NeubesitzerInnen zufrieden mit ihren Hunden. Einer der Hunde wurde aufgrund von Problemen mit dem Alleinbleiben zurückgegeben, im Wesenstest im Tierheim wurde er als problemlos beim Alleinbleiben eingeschätzt. Die Voraussagekraft für Reaktionen auf Menschen war unzureichend. Obwohl 33 (von 236) Hunde, aufgrund dessen durchgefallen sind, dass die starke Angst im Wesenstest gezeigt hatten, wurden zwölf der 45 vermittelten Hunde bei der BesitzerInnenbefragung dennoch als sehr ängstlich beschrieben und waren somit wohl falsch-negativ. Obwohl 20 (von 236) Hunde, die Aggression gegen andere Hunde im Wesenstest gezeigt hatten, aufgrund dessen durchgefallen sind, wurden zehn der 45 vermittelten Hunde bei der BesitzerInnenbefragung als aggressiv gegen mindestens einen anderen Hund beschrieben und waren somit vermutlich falsch-negativ. Obwohl elf (von 236) Hunde, die Fluchtverhalten oder exzessives Vokalisieren beim Alleinbleiben im Wesenstest gezeigt hatten, aufgrund dessen durchgefallen sind, gab es unter den 45 vermittelten Hunden 16 Hunde, welche nach der Vermittlung bei der BesitzerInnenbefragung während

des Alleinbleibens entweder aus dem Grundstück ausgebrochen sind oder exzessives Vokalisieren gezeigt haben und somit vermutlich falsch-negativ waren. Der negative prädiktive Wert dürfte demnach mangelhaft sein, da etwa ein Drittel der negativ-getesteten Hunde das unerwünschte Verhalten im neuen Zuhause dennoch zeigte. Die anderen statistischen Validitätskriterien sind aufgrund des fehlenden Referenzwertes nicht beurteilbar. Durch die Euthanasie der positiv-getesteten Hunde kann dieser nicht errechnet werden.

Die Befragung der NeubesitzerInnen durch CLAY et al. (2020a) erfolgte durch 36 Multiple-Choice-Fragen, angelehnt an die Messparameter des Wesenstests der RSPCA. Nur 123 NeubesitzerInnen nahmen teil, von denen drei Personen den Hund zum Befragungszeitpunkt nicht mehr in Besitz hatten. Lediglich die Verhaltensweisen *Freundlichkeit* und *Angst* konnten signifikant vorausgesagt werden, wohingegen Verhaltensprobleme, wie beispielsweise *Aggression*, *Futterverteidigung* und *Trennungsangst* nicht zuverlässig vorausgesagt werden konnten. Freundliche/soziale Verhaltensweisen im Wesenstest waren korreliert mit „*spielerischer/ freundlicher Art im Umgang mit BesitzerIn/Kindern/Fremden/Hunden*“. Ängstliche Reaktionen im Wesenstest waren korreliert mit „*der Hund geht auf Fremde/ Kinder/ BesitzerIn/ fremde Hunde auf besorgniserregende Art zu (Aggression)*“/ „*der Hund weicht vor BesitzerIn/Kindern zurück auf besorgniserregende Art*“. Alle anderen Verhaltensweisen im Wesenstest waren nicht mit den Reaktionen im neuen Zuhause korreliert. Aufgrund des Ausschlusses von Hunden mit aggressivem, besitzergreifendem oder trennungsangstbedingtem Problemverhalten aus der Vermittlung konnten auch in dieser Studie keine ausreichenden Angaben zu den statistischen Validitätskriterien Sensitivität und Spezifität gemacht werden. Es gab einige Falsch-Negative für Aggression, was auf einen unzureichenden negativen prädiktiven Wert für dieses Messparameter hinweist. Der Wert lag zwischen 2.88% (für Aggression gegen fremde Kinder) und 7.69% (Aggression gegen den bereits vorhandenen Hund). Diese Werte beziehen sich nur auf die Antworten der NeubesitzerInnen, es hatte vermutlich nicht jede/r NeubesitzerIn einen bereits vorhandenen Hund im Haushalt. Wäre dies der Fall gewesen, wäre der Prozentsatz vermutlich höher. Immerhin 9.2% der Hunde zeigten im neuen Zuhause Futteraggression, was ebenfalls auf einen unzureichenden negativen prädiktiven Wert hinweist. Die starke Korrelation beim Vorhandensein von Freundlichkeit und Angst deutet auf einen hohen positiven prädiktiven Wert für diese Messparameter hin.

Zur Untersuchung der prognostischen Validität des *B.A.R.K.* Wesenstests wurden 72 von 102 Hunden durch MORNEMENT et al. (2014) im neuen Zuhause zwei bis acht Monate nach Adoption durch ein Telefoninterview der NeubesitzerInnen reevaluiert. Eine Vorhersagekraft war lediglich für die Messparameter *Angst* und *Freundlichkeit* gegeben (signifikante Korrelationen zwischen *B.A.R.K.* und der BesitzerInnenbefragung) und ansonsten schwach. Bei den anderen Messparametern gab es keine signifikanten Korrelationen. Durch den Ausschluss von durch Aggressivität aufgefallenen Hunden wurde die Stichprobe ausschließlich durch nicht-aggressive Hunde gebildet, was das Ergebnis des Scorings von „*Freundlichkeit*“ beeinflusst haben könnte. Daher dürfte weiters die Beurteilung des wesentlichen Parameters *Aggression* in dieser Studie von vornherein unmöglich sein. Genaue statistische Angaben zur Validität sind der Studie nicht zu entnehmen.

In MORNEMENT et al. (2015) wurden die NeubesitzerInnen ebenfalls zwei bis acht Monate nach der Vermittlung via Telefoninterview befragt. Dazu wurde ein *B.A.R.K. Protocol Post Adoption Questionnaire* entwickelt, welcher am *B.A.R.K.* Wesenstest angelehnt ist. Es wurden 69 Fragen zum Auswahlprozess, zum Verhalten des Hundes im neuen Zuhause, zur Zufriedenheit mit dem neuen Hund und zur Adoptionserfahrung an sich gestellt. Ausgewertet wurden nur 73 Fragebögen, da ein Hund weitergegeben wurde. 82.4% der Hunde springen mindestens hin und wieder an Leuten hoch, 80.6% der Hunde ziehen mindestens hin und wieder stark an der Leine, 37.8% der Hunde reißen mindestens hin und wieder aus, 6.8% der Hunde verteidigen mindestens hin und wieder ihr Futter, 58.1% der Hunde haben mindestens hin und wieder etwas zerstört und 55.4% der Hunde waren mindestens hin und wieder gestresst oder nervös, wenn sie allein zuhause waren. Die Messparameter im *B.A.R.K. Protocol Post Adoption Questionnaire* zur Untersuchung der Voraussagekraft waren *ängstlicher/ unangemessener Toilettengang*, *Problemverhalten* und *Aggression*. Die Voraussagekraft war relativ schwach. Wurden die Messparameter *Angst* oder *Nervosität* im *B.A.R.K.* Wesenstest festgestellt, hat dies den Messparameter *ängstlichen/ unangemessenen Toilettengang im neuen Zuhause* vorausgesagt. Die Messparameter *Problemverhalten* und *Aggression* konnten nicht vorausgesagt werden. Die Punktescores im *B.A.R.K.* für *Angst* und *Nervosität* konnten signifikant voraussagen, wie oft die NeubesitzerInnen von *Angst gegenüber Fremden* berichten. Auch der Punktescore im *B.A.R.K.* für *Freundlichkeit* sagte voraus, wie oft Hunde BesucherInnen im neuen Zuhause freundlich begrüßten. Andere Verhaltensweisen nach der Adoption konnten nicht vorausgesagt werden.



Fast 25% der NeubesitzerInnen haben berichtet, dass ihr Hund geknurr, geschnappt oder versucht hat, eine Person zu beißen. Fast 75% der NeubesitzerInnen berichten, dass ihr Hund Verhalten zeigt, das sie gern ändern würden, wenn sie könnten. Dennoch haben über die Hälfte (56.8%) der NeubesitzerInnen berichtet, mit dem Verhalten ihres Hundes sehr zufrieden zu sein und 71.2% sagten, dass ihr Hund ihre Erwartungen erfüllt. Über 75% der NeubesitzerInnen würden in Zukunft wieder einen Hund adoptieren. Leider werden die Ergebnisse des *B.A.R.K.* Wesenstest nicht genau aufgeschlüsselt. Da in Australien Hunde, die im Tierheim positiv auf Aggression getestet wurden, gar nicht erst in die Vermittlung gehen, scheint die Rate der Falsch-Negativen erschreckend hoch zu sein, nachdem fast 25% der NeubesitzerInnen von aggressiven Vorfällen berichteten. Der negative prädiktive Wert dürfte daher gerade mal bei 75% liegen. In dieser Studie scheint *Angst* den höchsten positiven prädiktiven Wert zu haben.

Zur Untersuchung der prognostischen Validität des *Match-Up* Wesenstests wurden die NeubesitzerInnen durch MARDER et al. (2013) mindestens drei Monate (bis zu über zwei Jahren) nach der Adoption zum Auftreten von Futteraggression, wie sie die Futteraggression wahrnehmen und managen, sowie zu ihrer Zufriedenheit mit dem neuen Hund als Haustier befragt. Exkludiert wurden Hunde, deren NeubesitzerInnen sie nie während des Fressens unterbrochen haben, es sei denn, aufgrund der Tatsache, dass bereits Aggression gezeigt wurde in diesem Kontext. Insgesamt zeigten 29% (28 Hunde) Futteraggression im neuen Zuhause. Nur zwei dieser Hunde zeigten Futteraggression häufig oder immer, alle anderen Hunde nur selten. Nur fünf dieser Hunde zeigten erhebliche Futteraggression (Schnappen, Attackieren, Beißen). Fünfundfünfzig Prozent (11) der Hunde, die im Wesenstest Futteraggression zeigten, zeigten auch nach der Adoption Futteraggression. Zweiundzwanzig Prozent (17) der Hunde, die im Wesenstest keine Futteraggression zeigten, zeigten nach der Adoption dennoch Futteraggression. In dieser Studie wurden der negative prädiktive Wert und der positive prädiktive Wert untersucht. Im Vergleich zu anderen untersuchten Studien wurden die positiv getesteten Hunde ebenfalls vermittelt und somit ist eine bessere Beurteilung der statistischen Validitätskriterien möglich. Der positive prädiktive Wert liegt bei 55% und der negative prädiktive Wert liegt bei 78%. Die Sensitivität liegt bei 39.3% und die Spezifität liegt bei 87%.

In der Follow-Up-Studie zur Untersuchung der prognostischen Validität des *Ethotest* nach Adoption durch LUCIDI et al. (2005) wurden die Fähigkeit der sieben Hunde nach 14 Monaten PatientInnenkontakt analysiert. Die Ergebnisse des *Ethotest* haben sich bestätigt. Einer der beiden Hunde, die den *Ethotest* nicht bestanden haben, aber die Prüfung der *Delta Society*, wurde aus dem Arbeitsdienst exkludiert, da er in einer Situation zu grob reagiert hat. Hunde, die Schritt A nicht bestanden haben, sind eher unwahrscheinlich als Begleithund oder Therapiehund. Hunde, die Schritt B nicht bestanden haben, sollten noch einmal evaluiert werden und Hunde, die nur Schritt C nicht bestanden haben, benötigen spezielles Training, welches wünschenswert für Begleithunde und essentiell für Therapiehunde ist. Die Reevaluierung nach 14 Monaten ist allerdings weder auf das Leben im neuen Zuhause bezogen, noch ist die Stichprobengröße von sieben Hunden ausreichend.

STICCO et al. (2011) haben in Folge zur Evaluierung der prognostischen Validität des *Ethotest* die vermittelten zwölf Hunde für ein Jahr lang begleitet. Die/der PrüferIn, die/den der Hund kennt, besuchte den Hund nach zwei Wochen im neuen Zuhause. Anschließend wurden die NeubesitzerInnen alle zwei Monate telefonisch kontaktiert und zu ungewünschten Verhaltensweisen befragt. Keiner der zwölf Hunde zeigte im neuen Zuhause ungewollte Verhaltensweisen wie *Verunreinigung des Hauses*, *Hochspringen*, *Probleme beim Alleinbleiben* oder *Aggressivität*. Die Erwartungen der NeubesitzerInnen wurden erfüllt. Leider wurde einer der Hunde dem Neubesitzer aufgrund von Misshandlung abgenommen und jener Hund, welcher im *Ethotest* unter dem Mindestpunktescore lag, wurde ins Tierheim zurückgegeben. Die AutorInnen geben eine Sensitivität von 100% an, da alle adoptierten Hunde gewünschtes Verhalten zeigten und vorher positiv auf Vermittelbarkeit getestet wurden. Im Verständnis des bisherigen Vorgehens in dieser Arbeit würde dies aber eher den positiven prädiktiven Wert (der abgefragten Verhaltensweisen) beschreiben, da unklar ist, wie viele Hunde insgesamt tatsächlich positiv für das Zeigen von ausschließlich gewünschtem Verhalten sind. Für eine ausreichende Beurteilung der statistischen Validitätskriterien wären vollständige Informationen zu allen getesteten Hunden notwendig. Der Hund, welcher misshandelt und konfisziert wurde, zeigte keinerlei Zeichen von Aggression, sondern eher eine Art Depression, was von den AutorInnen als Beweis für die starke Vorhersagekraft des *Ethotest* spricht. Der Hund, welcher ins Tierheim zurückgegeben wurde, wurde ohnehin als unpassend für die Vermittlung bewertet, da sein Punktescore unter dem Mindestpunktescore lag. Auch dies sehen die AutorInnen als Bestätigung der Validität.

In Tabelle 33 findet sich eine Übersicht der prognostischen Validität. Die Beurteilung ist nur möglich, wenn eine Follow-Up-Studie im neuen Zuhause stattfand. Die meisten AutorInnen haben Sensitivität, Spezifität und die prädiktiven Werte angewendet. MORNEMENT et al. (2014; 2015) haben jedoch die Korrelation (anhand des *Korrelationskoeffizient nach Pearson*) zwischen dem Wesenstest und der Befragung der NeubesitzerInnen gemessen, daher ist die Vergleichbarkeit an dieser Stelle eingeschränkt.

Tab. 33: Übersicht über prognostische Validität von Wesenstests

Studie	Messparameter								
	Aggression gegenüber Menschen	Angst/ Furcht	Artgenossen	Trennungsangst	Trainierbarkeit/ Gehorsam	Erregung/ Aufregung	Freundlichkeit	Ressourcen-verteidigung/ Futteraggression	Vermittelbarkeit
Van der Borg et al., 1991	<i>Aggression ggü. Erwachsenen:</i> Sensitivität 82% NPW 85%		<i>Aggression ggü. Hunden:</i> Sensitivität 83% NPW 95%	Sensitivität 83% NPW: 92%	<i>Ungehorsam:</i> Sensitivität 84% NPW 91%				
Christensen et al., 2007 (Assess-A-Pet)	<i>ggü. BesitzerIn:</i> NPW 95.5%  <i>territorial:</i> NPW 48.5%  <i>Beuteaggression:</i> NPW 84%		<i>Aggression ggü. Hunden:</i> NPW 83%					NPW: 95.5%	
Bollen et Horowitz, 2008 (Assess-A-Pet)	Sensitivität 46-53% Spezifität 87-89% NPW 92%								
McGuire et al., 2020 (Assess-A-Pet)								<i>Spielzeug/ Knochen:</i> NPW: 89%  <i>Futter:</i> Sensitivität: 67% NPW 97.5%	

Sticco et al., 2011 (Ethotest)									Sensitivität 100%
Clay et al., 2020 (a) (Wesenstest der RSPCA)								<i>Futter:</i> NPW 91% <i>Spielzeug:</i> NPW 96%	
Mornement et al., 2014 (B.A.R.K.)		<i>Furcht:</i> konnte nicht signifikant vorausgesagt werden  <i>Angst:</i> mäßiger Zusammenhang r = 0.42 <sup>22</sup> P<0.01)			konnte nicht signifikant vorausgesagt werden	konnte nicht signifikant vorausgesagt werden	mäßiger Zusammenhang r = 0.49 (P<0.01)		
Mornement et al., 2015 (B.A.R.K.)	NPW 75%  keine signifikante Übereinstimmung	<i>Fremde:</i>  signifikante Übereinstimmung					<i>BesucherInnen:</i>  signifikante Übereinstimmung	<i>Futter:</i> NPW 93%	
Marder et al., 2013 (Match-Up)								<i>Futter:</i> PPW 55% NPW 78% Sensitivität 39% Spezifität 87%	

22 *Korrelationskoeffizient nach Pearson:* 0.0-0.2 kein bis geringer Zusammenhang, 0.21-0.5 schwacher bis mäßiger Zusammenhang, 0.51-0.8 deutlicher Zusammenhang, 0.81-1.0 hoher bis perfekter Zusammenhang (<https://www.medistat.de/glossar/korrelation-assoziaton/korrelationskoeffizient-nach-pearson/>, 2021)

### 3.4 Verhaltensbeurteilung durch Ethogramme

Unter einem Ethogramm versteht man die Beschreibung und Auflistung aller Verhaltensweisen, die bei einer Tierart unter natürlichen Bedingungen vorkommen, um ein möglichst vollständiges und vergleichbares Verhaltensrepertoire zu erhalten<sup>23</sup>.

Im Bezug auf die Vermittelbarkeit von Hunden kann der Unterschied zum Wesenstest darin bestehen, dass die Grundlage eines Ethogramms nicht in der Schaffung einer einmaligen expliziten Testsituation besteht, sondern in der langfristigen Beobachtung von Verhaltensweisen des jeweiligen Hundes im Tierheimalltag. Die meisten Studien hierzu wurden allerdings weniger mit dem Ziel der Evaluierung der Vermittelbarkeit, sondern eher zur Evaluierung der Lebensqualität der Hunde in der Tierheimsituation vorgenommen. Aus diesem Grund konnten nur drei Ethogramme untersucht werden.

#### 3.4.1 Ethogramme in der Praxis

##### 3.4.1.1 Ethogramm nach DE PALMA et al., 2005

Im Jahr 2005 haben DE PALMA et al. über die Fokus-Tier-Methode<sup>24</sup> und die ereigniszentrierte Beobachtung<sup>25</sup> 74 Tierheimhunde während eines Spaziergangs in einer neuen Umgebung beobachtet und in einem Ethogramm mit über 100 Verhaltensmustern erfasst. Die Hunde der Studie waren zum Zeitpunkt des Testes zwischen zwei und zwölf Monaten im Tierheim, welches ein sogenanntes *no-kill-shelter* ist. Das Ethogramm beinhaltet neun Beobachtungen für insgesamt 4,5 Stunden:

- 110 mögliche Verhaltensmuster im Zwinger, in dem der Hund mit einem zweiten Hund lebt, jeweils innerhalb von drei Beobachtungsperioden von 30 Minuten an drei unterschiedlichen Tagen

<sup>23</sup> [www.spektrum.de/lexikon/biologie](http://www.spektrum.de/lexikon/biologie)

<sup>24</sup> Die Fokus-Tier-Methode ist eine von vier Verfahren der Verfahrensbeobachtung, bei dem alle Vorkommnisse bestimmter Verhaltenskategorien eines Individuums für eine bestimmte Zeitspanne aufgezeichnet werden ([www.spektrum.de/lexikon/biologie](http://www.spektrum.de/lexikon/biologie)).

<sup>25</sup> Die ereigniszentrierte Beobachtung ist eine von vier Verfahren der Verhaltensbeobachtung, bei dem jedes Auftreten einer bestimmten Verhaltensweise in der gesamten beobachteten Gruppe erfasst wird ([www.spektrum.de/lexikon/biologie](http://www.spektrum.de/lexikon/biologie)).

- 115 mögliche Verhaltensmuster im Auslauf für zwei Beobachtungsperioden á 30 Minuten, allein mit der/dem BeobachterIn (ohne Interaktion)
- 25 mögliche Verhaltensmuster im Auslauf für zwei Beobachtungsperioden während der Wesenstest stattfindet, in dem die/der BeobachterIn und ein zweiter Hund involviert ist – es soll die Unabhängigkeit vom Menschen getestet werden
- die Hunde wurden weiters zweimal spazieren geführt, wobei verschiedene Reaktionen getestet wurden, jeweils für 30 Minuten.

Ziel der Studie war die Fähigkeit, durch die ereigniszentrierten Beobachtungen auf die Persönlichkeit des Hundes zu schließen, was teilweise möglich war. Es wurden keine Referenzwerte nach der Vermittlung durchgeführt, da der Fokus hier eher auf Hunden liegt, die langfristig im Tierheim leben.

#### **3.4.1.2. Ethogramm nach GOOLD et NEWBERRY, 2017**

In dieser Studie wurden longitudinale<sup>26</sup> Beobachtungsdaten erfasst unter Verwendung des generellen Modells von Verhaltensreaktionsnormen.

Die Varianz wurde untergliedert in:

- Persönlichkeit (Unterschiede zwischen den Individuen im durchschnittlichen Verhalten)
- Formbarkeit (individuelle Unterschiede bei Verhaltensänderungen)
- Vorhersagekraft (Unterschiede in der Restvariabilität des einzelnen Individuums).

Untersucht wurden alle Interaktionen von 3263 Hunden mit fremden Personen im Test, für einen Monat lang nach Aufnahme im Tierheim. Es wurden alle beobachteten Interaktionen notiert, wenn ein/e MitarbeiterIn sich einem fremden Hund genähert hat oder wenn ein/e MitarbeiterIn die Annäherung einer fremden Person an den Hund beobachtet hat. Verschiedene MitarbeiterInnen konnten ihre Beobachtungen eintragen.

Das Ethogramm beinhaltet 13 Verhaltensweisen (*13-code Ethogramm*) inklusive ihrer Definition und einer subjektiven Kennzeichnung entsprechend ihrer Vermittelbarkeit (grün –

---

<sup>26</sup> *longitudinal* bedeutet im Gegensatz zu *transversal*, dass die Individuen mehrfach beobachtet werden ([www.spektrum.de/lexikon/biologie](http://www.spektrum.de/lexikon/biologie)).

gelb – rot). Grüne Verhaltensweisen bedeuten keine Probleme für die Vermittlung, gelbe Verhaltensweisen bedeuten, dass Training empfohlen wird, der Hund aber keine Gefahr für Menschen und andere Hunde darstellt. Rote Verhaltensweisen bedeuten, dass der Hund eine Gefahr für Menschen und/oder andere Hunde darstellen kann und Verhaltensmodifikation zwingend notwendig ist. Eingetragen wurde immer der höchste Code, der während einer Beobachtung identifiziert wurde. Pro Tag wurde maximal eine Eintragung vorgenommen. Im Durchschnitt wurden pro Hund 5.9 Beobachtungen eingetragen.

#### **3.4.1.3 Ethogramm nach CLAY et al., 2019**

In der Studie von CLAY et al. (2019) wird untersucht, ob die Beobachtung von Hunden in ihren Zwingern frühe Hinweise auf Problemverhalten bieten kann. Hierzu wurden 38 Hunde in einem Tierheim in den ersten fünf Tagen ihres Tierheimaufenthaltes beobachtet. Die Beobachtung erfolgte für 60 Minuten täglich zur selben Uhrzeit via Kamera. Die Ergebnisse dieser Beobachtungen wurden mit den Ergebnissen eines Wesenstests verglichen, welcher fünf Tage nach der Aufnahme im Tierheim stattfand.

#### **3.4.2 Vergleichende Übersicht aller Ethogramme**

In der folgenden Tabelle (Tab. 34) findet sich eine Übersicht über die gemessenen diagnostischen Kriterien, aufgeschlüsselt auf die einzelnen Studien. Leider misst keine der Studien alle relevanten Kriterien.



Tab. 34: Übersicht über die untersuchten diagnostischen Kriterien in den Studien

Studie	Anzahl Messparameter	Anzahl Verhaltensweisen	Objektivität				Reliabilität			Validität	
			standardisierter Ablauf	Anzahl BeurteilerInnen	Objektivität	Praktikabilität	Interne Konsistenz	Test-Retest-Reliabilität	Inter-Rater-Reliabilität	Intra-Rater-Reliabilität	Inhaltsvalidität
De Palma et al., 2005	5	16	✓	2			✓		✓		
Goold et Newberry, 2017	1	13	-	n.d.				✓	✓		
Clay et al., 2019	5	48	✓	n.d.							

In der folgenden Tabelle (Tab. 35) findet sich eine Übersicht der Messparameter, welche in den untersuchten Studien betrachtet wurden.

Tab. 35: Übersicht über die untersuchten Messparameter

Studie	Messparameter						
	Aggression ggü. Menschen	Angst/ Furcht	Artgenossen	Erregung/ Nervosität	Freundlichkeit/ Umgänglichkeit	Vermittelbarkeit	weitere
De Palma et al., 2005	+		Dominanz/ Angst		+		+
Goold et Newberry, 2017						+	
Clay et al., 2019	+	+		+			

### **3.4.3 Diagnostische Testkriterien der Ethogramme**

#### **3.4.3.1 Objektivität von Ethogrammen**

Das Hauptproblem der Ethogramme dürfte die Praktikabilität sein, da die Auswertung von Ethogrammen sehr zeitintensiv ist. Die Verhaltensbeurteilung durch DE PALMA et al. (2005) nimmt eine Dauer von 4,5 Stunden in Anspruch. Die Beobachtung von CLAY et al. (2019) dauert fünf Stunden.

Die Beurteilung durch GOOLD et NEWBERRY ist nicht standardisiert.

Die Stimuli, auf die die Reaktionen beobachtet wurden, waren spontane Interaktionen und weisen daher keine Standardisierung auf (GOOLD et NEWBERRY, 2017). Dies wird aber dadurch begründet, dass es repräsentativer für das Verhalten außerhalb des Tierheims ist.

#### **3.4.3.2 Reliabilität von Ethogrammen**

Es wurde in DE PALMA et al. (2005) die Inter-Rater-Reliabilität gemessen zwischen den Autorinnen, welche die Beobachtungen durchführten. Sie lag zwischen 0.82 und 0.89 und ist somit akzeptabel. Die Inter-Rater-Reliabilität wurde auch durch GOOLD et NEWBERRY (2017) untersucht und für signifikant befunden. Die durchschnittliche Übereinstimmung für die Bewertung eines Videos, welches ein Beispiel für nicht-aggressive Reaktion auf Menschen lag bei 0.75 und bei einem Beispiel für aggressive Reaktion auf Menschen bei 0.77.

Die Wiederholpräzision des Verhaltens nahm bei GOOLD et NEWBERRY (2017) innerhalb der ersten Woche Aufenthalt im Tierheim zu.

Die interne Konsistenz der Verhaltenskategorien in den fünf Messparametern bei DE PALMA et al. (2005) wurde getestet, dabei waren 12 von 16 Kategorien konsistent für die Messparameter.

CLAY et al. (2019) haben keine Untersuchungen zur Reliabilität vorgenommen.

#### **3.4.3.3 Validität von Ethogrammen**

Keine der Studien untersucht die Vorhersagekraft von Ethogrammen unter Verwendung des Referenzwertes „*Verhalten nach Adoption*“, in keiner Studie wurde ein Follow-Up durchgeführt.

Die Verhaltensentwicklung im Beobachtungszeitraum spiegelte sich durchschnittlich in einer Zunahme des Verhaltens *Freundlichkeit* wider (GOOLD et NEWBERRY, 2017). *Aggression* wurde selten beobachtet und die Wahrscheinlichkeit für die Beobachtung von „roten Verhaltenscodes“ nahm im Laufe der Tage ab. Die Studie ergab aber, dass es sehr unsicher ist, die individuelle Persönlichkeit von Hunden aufgrund einiger Beobachtungen zu klassifizieren. Einzelne Hunde zeigten im Gegensatz zum Durchschnitt eine Verminderung der Umgänglichkeit. Einige Hunde zeigten starke Veränderungen innerhalb der ersten Tage, andere Hunde zeigten nur sehr langsam Verhaltensveränderungen. Auch die Evaluierung der *Formbarkeit* unterliegt großer Unsicherheit. Die sozialsten Hunde am Tag der Aufnahme im Tierheim waren nicht automatisch die sozialsten Hunde einige Tage später. Die Vorhersagekraft war ebenfalls stark gestreut, was für variierende Level von Verhaltenskonsistenz bei Hunden spricht. Eine stärkere Formbarkeit war assoziiert mit einer niedrigeren Vorhersagekraft. Zwischen Persönlichkeit und Vorhersagekraft wurden keine Korrelationen gefunden.

Die Ergebnisse der Beobachtungen durch CLAY et al. (2019) wurden mit den Ergebnissen eines Wesenstests verglichen, welcher fünf Tage nach der Aufnahme im Tierheim stattfand und somit als Referenzwert galt. Die Messparameter entsprechen denen des Ethogramms. Es wurden die verschiedenen Verhaltensweisen im Ethogramm in ihrem Zusammenhang mit dem Ausgang des Wesenstests (Adoption/ Euthanasie) untersucht. Zwingerverhalten in Form von Angst, gezielter Furcht und Erregung war signifikant korreliert mit ebendiesem Verhalten im Wesenstest. Wurden einige Verhaltensweisen vermehrt beobachtet, wie beispielsweise Weinen, angespannte Körperhaltung oder Hecheln, hatten diese Hunde eine erhöhte Chance, den Wesenstest nicht zu bestehen. Es wurde ein Zusammenhang zwischen dem Verhalten im Zwinger in den ersten fünf Tagen und dem Verhalten im Wesenstest gefunden. Genaue Angaben zu statistischen Validitätskriterien können nicht entnommen werden.

## 4 Diskussion

Insgesamt wurden 19 verschiedene publizierte Verfahren zur Beurteilung der Vermittelbarkeit, des Aggressionspotentials und/oder des Charakters von Einzelhunden im Tierheim auf die Erfüllung wichtiger diagnostischer Testparameter untersucht. Darunter waren drei Fragebögen, 13 Wesenstests und drei Ethogramme.

Für die Untersuchung der Verfahren haben allerdings nur 34 Studien den Einschlusskriterien entsprochen und konnten damit in die Arbeit einfließen. Nur vier dieser 34 Studien konnten der Evidenzklasse *Silber* zugeordnet werden, alle weiteren Studien erreichten lediglich die Bronzекlasse.

Für keines der 19 Beurteilungsverfahren wurden im Rahmen der publizierten Studien alle relevanten Testparameter untersucht.

Diese Punkte sind klare Hinweise darauf, dass es weiterer fundierter wissenschaftlicher Studien benötigt und die wissenschaftliche Qualität dieser Arbeiten gesteigert werden muss, da die Folgen für die beurteilten Hunde schwerwiegend sein können.

Die Diskussion erfolgt ähnlich der Aufarbeitung der Ergebnisse ebenfalls in Unterkapiteln.

### 4.1 Fragebögen

Für keinen der drei Fragebögen (*C-BARQ*, *Fragebogen nach STEPHEN et LEDGER*, *Monash/ MCPQ-R*) wurden alle als relevant befundenen Testparameter in den vorhandenen publizierten Studien untersucht.

#### 4.1.1 Objektivität von Fragebögen

Bei der Frage, ob die untersuchten Testparameter erfüllt werden, taucht das größte Problem bereits hinsichtlich der Objektivität auf. Unter den AutorInnen herrscht Uneinigkeit darüber, ob abgebende BesitzerInnen korrekte Angaben machen, da diese wissen, dass der Fragebogen die Folgen für den jeweiligen Hund beeinflusst und daher bestimmte Verhaltensprobleme verschwiegen werden könnten.

SEGURSON et al. (2005; *C-BARQ*) bestätigen diese Diskrepanz; bei nur einem von sieben Messparametern (*Bemühen um Aufmerksamkeit*) ergab sich kein signifikanter Unterschied

zwischen den Angaben von zwei BesitzerInnengruppen (*vertraulich* vs. *nicht-vertraulich*). Während das *Bemühen um Aufmerksamkeit* brauchbare Antworten in Aufnahmefragebögen geben kann, dürfte dies beim Thema *Aggression* nicht der Fall sein. Nachdem aber gerade Aggression gegen Personen im Haushalt ein Hauptgrund für Hundebisse sein kann, sind hier die falschen Angaben der abgebenden Personen besonders besorgniserregend. Weiters ist dieser Punkt sehr ausschlaggebend für eine erfolgreiche Vermittlung und wäre daher wichtiger Bestandteil eines aussagekräftigen Fragebogens. Die Studienergebnisse geben den Beweis dafür, dass abgebende BesitzerInnen in manchen Aspekten des Fragebogens falsche Angaben machen und somit ist es unklar, ob BesitzerInnen, die von keinen Verhaltensproblemen berichten, tatsächlich einen problemfreien Hund abgeben oder befangene Antworten geben, um die Wahrscheinlichkeit auf dessen Adoption zu erhöhen. Die nicht-vertrauliche Gruppe hat in einigen Kategorien signifikant öfter mit „Abwesenheit des Verhaltens“ geantwortet, was diese Befangenheit widerspiegeln könnte.

Im Gegensatz zu diesen Erkenntnissen ergab die Studie von DUFFY et al. (2014) zum Fragebogen *C-BARQ* keinen einheitlichen Beweis dafür, dass abgebende BesitzerInnen unzuverlässige oder voreingenommene Antworten im Fragebogen geben, abhängig davon ob sie glaubten, dass diese Informationen an Tierheimpersonal weitergegeben oder zur Evaluierung der Vermittelbarkeit verwendet werden oder nicht. Die Stichprobengröße wurde größer gewählt als durch SEGURSON et al. ( $N = 438$  vs.  $N = 54$ ). Als weiterer möglicher Grund für diese Unterschiede wurde die Tatsache aufgeführt, dass SEGURSON et al. (2005) persönliche Interviews mit den abgebenden BesitzerInnen führten (ist der Studie selbst allerdings nicht zu entnehmen) anstatt von schriftlichen Befragungen wie in dieser Studie. Laut DUFFY et al. (2014) wird bestätigt, dass der Fragebogen im Großen und Ganzen geeignet ist zur Feststellung von Verhaltensproblemen bei Abgabehunden. Die Aussage von SEGURSON et al. (2005), dass abgebende BesitzerInnen unter gewissen Umständen Falschangaben machen, scheint widerlegt. Es stellt sich jedoch auch die Frage, ob Personen in einer direkten Befragung mehr Hemmungen haben, den Interviewer/ die Interviewerin anzulügen und dadurch die Diskrepanz der beiden Studien ebenfalls zu erklären ist.

Dennoch halte ich eine objektive Einschätzung durch abgebende BesitzerInnen für eher unrealistisch. Sollten Fragebögen etabliert sein, wird abgebenden BesitzerInnen sowohl klar sein, dass die Daten weiterverwendet werden für die Vermittlung. Auch wenn dieses Wissen

laut DUFFY et al. (2014; *C-BARQ*) keinen Einfluss auf die Angaben im Fragebogen hat, werden die abgebenden BesitzerInnen mindestens von ihren persönlichen Emotionen beeinflusst, schließlich handelt es sich um eine zumeist schwere Entscheidung nach einer oft langen Entscheidungsphase, die zum Verzicht auf ein in der Regel geliebtes Familienmitglied führte.

#### **4.1.2 Reliabilität von Fragebögen**

Keine Studie zur Anwendung von Fragebögen bei Abgabe im Tierheim befasste sich mit der Intra- und Inter-Rater-Reliabilität. Es stellt sich dabei die Frage der Sinnhaftigkeit dieser Messungen, da die VorbesitzerInnen meist zu einem einmaligen Zeitpunkt (bei Abgabe des Hundes im Tierheim) den Fragebogen ausfüllen und zu diesem Zeitpunkt auch nur sie in der Lage sind, den Hund zu beurteilen. Auch die Untersuchung der Test-Retest-Reliabilität ist bei abgebenden BesitzerInnen aufgrund der einmaligen Befragung zum Zeitpunkt der Abgabe im Tierheim kaum möglich.

##### **4.1.2.1 Test-Retest-Reliabilität von Fragebögen**

Die Test-Retest-Reliabilität bei zwei Befragungen der NeubesitzerInnen dürfte hoch sein (STEPHEN et LEDGER, 2007; *Fragebogen nach STEPHEN et LEDGER*). Die Veränderung mancher Verhaltensweisen des Hundes zwischen Woche zwei und Woche sechs spricht aber dennoch dafür, dass das Verhalten des Hundes sogar in einer weitestgehend stabilen Umgebung nicht konsistent ist. Außerdem taucht die Frage auf, ob eine hohe Test-Retest-Reliabilität bei NeubesitzerInnen gleichgesetzt werden kann mit der Befragung abgebender BesitzerInnen.

Ziel dieser Arbeit war es explizit, die Beurteilungsverfahren für die Population der Tierheimhunde zu betrachten. Dabei stellen sich natürlich aber auch weitere Fragen: Gibt es neben den bereits untersuchten Methoden andere Fragebögen (oder Wesenstests), welche bisher nur in einer anderen Population als der Tierheimhundepopulation untersucht wurden und dabei aber die wichtigsten Testkriterien erfüllten? Oder haben jene Evaluierungsverfahren, welche im Zuge der Diplomarbeit bereits untersucht wurden, bei Hunden in Privatbesitz die Testkriterien erfüllt? Aus diesem Grund werden in der Diskussion ergänzend die Ergebnisse von anderen Studien erfasst, um die diagnostischen Testparameter bei Hunden in Privatbesitz mit Tierheimhunden vergleichen zu können.

So wurde beispielsweise der *Monash Canine Personality Questionnaire* für Hunde in Privatbesitz auf seine Test-Retest-Reliabilität untersucht (LEY et al., 2009a). Hierfür hatte ein/e BesitzerIn den Fragebogen nach einer gewissen Zeit ein zweites Mal auszufüllen. Die Übereinstimmung war für alle fünf Messparameter statistisch signifikant. *Neurotizismus* erreichte die niedrigste (0.79) und *Motivation* die höchste Test-Retest-Reliabilität (0.93).

Auch JONES (2008) hat einen Fragebogen (*Dog Personality Questionnaire DPQ*) entwickelt und dessen Test-Retest-Reliabilität in einer Population von Hunden in Privatbesitz untersucht. Die TeilnehmerInnen wurden erneut rekrutiert, um den Fragebogen ein zweites Mal auszufüllen. Die Messwerte variieren nach Messparameter und Länge des Fragebogens zwischen 0.872 (*Aggression gegenüber Menschen* in der Kurzfassung) und 0.939 (*Angst* in der Langfassung) und sind somit gut.

#### **4.1.2.2 Inter-Rater-Reliabilität von Fragebögen**

Neben den Studien von SEGURSON et al. (2005) und DUFFY et al. (2014) wurden über 25 Studien zu *C-BARQ* veröffentlicht (WIENER et HASKELL, 2016), allerdings ebenfalls nicht im Zusammenhang mit der Abgabe im Tierheim. Die Inter-Rater-Reliabilität lag hierbei zwischen 0.53 und 0.9 und war somit sehr uneinheitlich.

Der *Monash Canine Personality Questionnaire* wurde für Hunde in Privatbesitz auf seine Inter-Rater-Reliabilität untersucht (LEY et al., 2009a). Hierfür wurden BesitzerInnenpaare angehalten, jeweils den eigenen Hund zu beurteilen. Die Übereinstimmung war für alle fünf Messparameter statistisch signifikant. *Trainingsfokus* erreichte die niedrigste (0.73) und *Extroversion* die höchste Inter-Rater-Reliabilität (0.86). Die Untersuchung der Inter-Rater-Reliabilität könnte daher im Tierheimkontext analog zu LEY et al. (2009a) vorgenommen werden, indem nach Möglichkeit sämtliche Familienmitglieder der abgebenden Familie den Fragebogen ausfüllen. Diese Untersuchung wäre sehr interessant, da eine niedrige Übereinstimmung für die Angabe falscher Daten sprechen könnte. Vertraulichkeit gegenüber den anderen BeurteilerInnen innerhalb einer Familie wäre allerdings unbedingt erforderlich, da, wie erwähnt, bei der Abgabe eines Hundes viele emotionale, aber auch zwischenmenschliche Faktoren eine Rolle spielen. Auch die Festlegung eines Mindestalters bei der Teilnahme an der Befragung wäre eine interessante Frage, da jüngere Familienmitglieder eine völlig andere Sicht auf eventuell vorliegende Probleme haben könnten.

Auch JONES (2008) hat für ihren Fragebogen (Dog Personality Questionnaire DPQ) die Inter-Rater-Reliabilität in einer Population von Hunden in Privatbesitz untersucht. Auch hier wurden HundebesitzerInnenpaare aufgefordert, den Fragebogen auszufüllen, ohne sich miteinander abzusprechen. Die Inter-Rater-Reliabilität variiert nach Messparameter und Länge des Fragebogens zwischen 0.659 (*Aggression gegenüber Menschen* in der Langfassung) und 0.786 (*Aktivität/ Aufregung* in der Langfassung) und ist somit moderat bis gut.

In einer Studie (MIRKÒ et al., 2013) zum *Dog Personality Questionnaire (DPQ)* nach MIRKÒ et al. wurden die Berichte der BesitzerInnen (von Hunden in Privatbesitz) verglichen mit den Auswertungen der Videos eines Wesenstests (*FIDO Personality test for dogs*) dieser Hunde. Der *DPQ*-Fragebogen umfasst die Messparameter *Umgänglichkeit gegenüber Fremden*, *Aktivität*, *Aggression* und *Trainierbarkeit*. Die Auswertung erfolgte durch drei Gruppen (TrainerInnen von nicht-inkludierten Hunden, BesitzerInnen von nicht-inkludierten Hunden und Nicht-HundebesitzerInnen). Die paarweisen Vergleiche waren signifikant korreliert, außer zwischen den BesitzerInnen und allen anderen Gruppen bezüglich *Trainierbarkeit*. Die Korrelationen variierten abhängig von Messparameter und BeurteilerInnenpaar zwischen 0.38 (TrainerInnen vs. BesitzerInnen nicht-inkludierter Hunde im Messparameter *Trainierbarkeit*) und 0.97 (Nicht-HundebesitzerInnen vs. BesitzerInnen nicht-inkludierter Hunde im Messparameter *Aktivität*).

#### **4.1.2.3 Intra-Rater-Reliabilität von Fragebögen**

Denkbar wäre an dieser Stelle alternativ eine Befragung durch die jeweiligen HundetrainerInnen im Vorfeld, wenn die Abgabe im Tierheim im Raum steht und das zweite Mal bei der Abgabe selbst, um diesen Punkt evaluieren zu können. Aber aufgrund der emotionalen Einflüsse, die eine mögliche und später tatsächliche Abgabe des Hundes mit sich bringen, könnten diese Ergebnisse auch bei vermeintlich objektiveren Personen wie HundetrainerInnen verfälscht werden.

Untersuchungen der Intra-Rater-Reliabilität von Fragebögen liegen auch für Hunde in Privatbesitz nicht vor.



### 4.1.3 Validität von Fragebögen

Die Vorhersagekraft der Fragebögen variiert stark; sowohl nach den gemessenen Messparametern als auch nach dem verwendeten Fragebogen.

Der Zeitpunkt der Befragung der NeubesitzerInnen könnte hier eine große Rolle spielen. So wurde zum Beispiel in vier Punkten des Fragebogens nach STEPHEN et LEDGER (2007) eine signifikante Zunahme von Problemverhalten von Woche zwei zu Woche sechs post Adoption festgestellt. Es scheint daher sinnvoll, BesitzerInnenbefragungen frühestens nach sechs Wochen nach Adoption durchzuführen.

Im Vergleich der vertraulichen Gruppe mit einer Gruppe von HundebesitzerInnen, welche ihre Hunde zum Zeitpunkt der Evaluierung nicht im Tierheim abgeben wollten, wichen sechs Messparameter signifikant voneinander ab (SEGURSON et al. 2005; C-BARQ). Abgabehunde hatten signifikant höhere Punktescores bezüglich *Aggression gegenüber dem/der BesitzerIn*, *Aggression gegenüber Fremden*, *Angst gegenüber Fremden*, *Aggression gegenüber Hunden*, *nicht-soziale Angst* und *Trennungsverhalten*. Diese Erkenntnis bestätigt, dass Hunde mit Verhaltensproblemen ein höheres Risiko haben, im Tierheim abgegeben zu werden. Man kann hier den Unterschied zwischen der Prävalenz und der Vortestwahrscheinlichkeit erkennen. Die Prävalenz der o.g. Parameter, die offensichtlich ausschlaggebend dafür sind, einen Hund abzugeben war in der abgebenden Gruppe höher. Allerdings sind häufig auch Zeitmangel, Platzmangel, unrealistische Erwartungen der abgebenden BesitzerInnen, sowie deren Mangel an Wissen über die Bedürfnisse und das Verhalten von Hunden an der Entstehung von Verhaltensproblemen (wie z.B. Trennungsangst) mitbeteiligt und das ursächliche Problem ist somit unklar. So ist Trennungsangst beispielsweise häufiger bei Hunden aus dem Tierheim als bei Hunden aus anderen Quellen. Es stellt sich die Frage, ob die Hunde mit Trennungsangst häufiger im Tierheim abgegeben werden oder ob Hunde im Tierheim durch die Umstände vor Ort Trennungsangst entwickeln. Die signifikant größere Prävalenz für Trennungsangst in dieser Studie bei abgebenden BesitzerInnen im Vergleich zu nicht-abgebenden BesitzerInnen spricht dafür, dass Trennungsangst ein bereits vorher existierendes Verhaltensproblem ist. Auch STEPHEN et LEDGER (2007) haben in ihrem Fragebogen bezüglich *Furcht beim Alleinbleiben* ebenfalls deutliche statistisch signifikante Übereinstimmung zwischen Vor- und NeubesitzerInnen gefunden. Diese war auch bei beiden Befragungen der NeubesitzerInnen

(zwei und sechs Wochen nach Adoption) deutlich. Dem widersprechen wiederum DUFFY et al. (2014), welche im C-BARQ Fragebogen keine Korrelation beim Alleinbleiben zwischen Vor- und NeubesitzerInnen gefunden haben. Der Zeitpunkt der Follow-Up-Studie scheint als Grund der Abweichungen unwahrscheinlich, da sechs Wochen (STEPHEN et LEDGER, 2007) und zwei Monate (DUFFY et al., 2014) nicht weit auseinander liegen.

Fragebögen haben bezüglich *Aggression gegenüber fremden Personen* eine moderate bis deutliche signifikante Korrelation erreicht (STEPHEN et LEDGER, 2007; DUFFY et al., 2014; C-BARQ), bezüglich *Aggression gegenüber den BesitzerInnen* jedoch lediglich eine schwache Korrelation (DUFFY et al., 2014; C-BARQ), auch hier stellt sich jedoch die Frage der Befangenheit der ausfüllenden BesitzerInnen. Demographische Unterschiede könnten hier ebenfalls eine Rolle spielen. Aufgrund der Tatsache, dass Hunde mit einer Vorgeschichte von Aggression im Vereinigten Königreich kaum Vermittlungschancen haben oder durch das Tierheim gar nicht erst aufgenommen werden, könnte man auch beim *Fragebogen nach STEPHEN und LEDGER (2007)* denken, dass VorbesitzerInnen von diesem Messparameter zu einem verminderten Grad berichten. Die Korrelationen für den Messparameter *Aggression* waren aber bei STEPHEN und LEDGER im Vergleich zu den anderen Messparametern nicht schwächer, sodass dies ebenfalls keine Erklärung darzustellen scheint.

Die Aussagekraft von Fragebögen zum Messparameter *Angst* ist mit den vorliegenden Daten von nur vier Studien, wovon nur zwei *Furcht/Angst* beurteilt haben, nur bedingt beurteilbar. Einige AutorInnen haben zusätzlich unterschieden in Angst im Allgemeinen (*anxiety*) und Furcht vor bestimmten Reizen (*fear*), so beispielsweise STEPHEN et LEDGER (2007). Während in den C-BARQ-Fragebögen laut DUFFY et al. (2014) keine Korrelationen bezüglich dieser Messparameter zwischen Befragungen der Vor- und NeubesitzerInnen festgestellt werden konnten, war im *Fragebogen nach STEPHEN et LEDGER (2007)* eine moderate statistisch signifikante Korrelation nur bezüglich *Angst gegenüber dem/der TierärztIn* vorhanden – ein Aspekt, der meiner Meinung nach nicht existenziell für eine Vermittelbarkeit relevant ist.

SEGURSON et al. (2005) haben im C-BARQ Fragebogen *Angst und Aggression gegenüber Artgenossen* erfasst, DUFFY et al. (2014) haben später den C-BARQ um *Rivalität gegenüber Artgenossen* erweitert. Die Korrelation zwischen der Befragung von Vor- und

NeubesitzerInnen war bei SEGURSON et al. schwach, bei DUFFY et al. moderat für *Aggression gegenüber Artgenossen*.

Ressourcenverteidigung wurde nur durch MCGUIRE et al. (2020; C-BARQ) ermittelt und erlangte einen hohen negativen prädiktiven Wert (93.6%), Futter wurde dabei allerdings nicht getrennt untersucht. Der positive prädiktive Wert (38.5%) ist unzureichend aussagekräftig, die Berücksichtigung bei Auftreten dieses Punktes im Sinne einer negativen Beurteilung erscheint daher äußerst fragwürdig.

Die Sensitivität des C-BARQ Fragebogens (SEGURSON et al., 2005) bezüglich der Messparameter *Aggression gegenüber BesitzerIn/ Familienmitglieder* und *Angst gegenüber Fremden* scheint niedrig zu sein, da BesitzerInnen das Vorhandensein von diesen Problemverhalten nicht angeben und sie somit nicht durch den Fragebogen erfasst werden, wodurch sich wiederum viele Falsch-Negative ergeben. Der negative prädiktive Wert bezüglich der o.g. Messparameter scheint daher ebenfalls eher niedrig zu sein. Die Spezifität des Fragebogens sollte in allen Messparametern eher hoch sein, da die abgebenden BesitzerInnen die Hunde am besten kennen und die Abwesenheit von Problemverhalten richtig angeben können. Hat ein Hund demnach ein bestimmtes Verhaltensproblem nicht, haben abgebende BesitzerInnen vermutlich wenig Grund, dem Hund dieses Verhaltensproblem dennoch nachzusagen, daher schätze ich die Zahl an Falsch-Positiven als eher gering ein. Der positive prädiktive Wert sollte aus eben diesem Grund in allen Messparametern eher hoch sein, da Problemverhalten, sollte es von den abgebenden BesitzerInnen angegeben worden sein, vermutlich auch tatsächlich beobachtet wurde im täglichen Zusammenleben mit dem Hund. Geht man allerdings davon aus, dass abgebende BesitzerInnen nicht nur Problemverhalten verschweigen ("*under-reporting*"), sondern im Allgemeinen im Fragebogen unehrlich sind, ist mit niedrigen Werten bei allen prognostischen Parametern zu rechnen. Die Referenzwerte wurden in dieser Studie nicht erfasst, das heißt, die tatsächliche Anwesenheit/ Abwesenheit bestimmter Verhaltensprobleme wurde nicht überprüft, was die genaue Sensitivität und Spezifität nicht messbar macht.

Wenn man davon ausgeht, dass *Bemühen um Aufmerksamkeit* im C-BARQ durch die VorbesitzerInnen richtig angegeben wird (SEGURSON et al., 2005), wird die Vorhersagekraft dieses Parameters durch DUFFY et al. (2014) dennoch widerlegt, da die Übereinstimmung dieses Parameters zwischen Vor- und NeubesitzerInnen nicht signifikant ist. Fasst man die

Aussagen der Studie von DUFFY et al. (2014) zum C-BARQ zusammen, kann man daraus schließen, dass lediglich folgende Messparameter im Fragebogen eine signifikante Vorhersagekraft aufweisen: *Aggression gegenüber Fremden*, *Urinabsatz beim Alleinbleiben* und *Kauen von unangebrachten Objekten*. Für *Aggression gegenüber Fremden* könnte dieses Ergebnis, entgegen der Annahme von SEGURSON et al. (2005), für korrekte Angaben durch die VorbesitzerInnen sprechen. Dennoch bleibt die Frage offen, ob abgebende BesitzerInnen *Aggression gegenüber Fremden* vielleicht ehrlicher angeben als *Aggression gegenüber BesitzerIn/ Familienmitglieder*. Die generell großen Abweichungen im Vergleich der C-BARQ-Ergebnisse (DUFFY et al., 2014) mit der Befragung der NeubesitzerInnen scheinen dafür zu sprechen, dass die Hunde in den unterschiedlichen Situationen sehr unterschiedliche Reaktionen zeigen, was den Vergleich von Verhaltens-evaluierungen zwischen Vor- und NeubesitzerInnen in Frage stellt und dadurch verständlicherweise einen schwachen Vorhersagewert bedingt.

Neben den Studien von SEGURSON et al. (2005) und DUFFY et al. (2014) wurden über 25 Studien zu C-BARQ veröffentlicht (WIENER et HASKELL, 2016), allerdings nicht im Zusammenhang mit der Abgabe im Tierheim. So wurde erstmals in der initialen Studie (SERPELL et HSU, 2001) die Kriteriumsvalidität des C-BARQ gemessen, indem bei der Auswahl von Assistenzhunden die Ergebnisse des Fragebogens mit einem Referenzwert verglichen wurde. Diesen Referenzwert bildeten die betriebseigenen Kriterien einer Assistenzhundeschule *„zur Ablehnung von Hunden aus verhaltensbedingten Gründen“*. Fast alle erwarteten Übereinstimmungen zwischen Fragebogen und Referenzwert waren signifikant oder grenzwertig signifikant. Der Fragebogen wurde damit als valide befunden. In der Folgestudie (HSU et SERPELL, 2003) wurde der C-BARQ Fragebogen bei Hunden angewandt, welche durch HundeverhaltensexpertInnen aufgrund von Verhaltensproblemen untersucht wurden. Die Diagnosen dieser ExpertInnen dienten in dieser Studie als Referenzwert für die Validität des Fragebogens. Die Ergebnisse des Fragebogens korrelierten in allen sieben Verhaltenskategorien (*Aggression gegenüber BesitzerInnen/ Fremden/ Hunden*, *Angst vor Fremden/ Hunden/ Geräuschen/ Gewittern*, *Trennungsangst*, *Bemühen um Aufmerksamkeit*) signifikant mit den Befunden der HundeverhaltensexpertInnen. Eine weitere Studie (BARNARD et al., 2012) verglich die Ergebnisse des C-BARQ mit den Ergebnissen eines kurzen Wesenstests (bestehend aus vier Einzeltests).

Signifikante Korrelationen wurden gefunden bezüglich *Aggression gegenüber Kindern, Angst/ Aggression gegenüber Fremden, Angst/ Aggression gegenüber Artgenossen*.

JONES (2008) hat für ihren Fragebogen (*Dog Personality Questionnaire DPQ*) die prognostische Validität in einer Population von Hunden in Privatbesitz untersucht. Als Referenzwert wurde ein Wesenstest verwendet. Aggressive Verhaltensweisen im Wesenstest haben signifikant korreliert mit Aggression im Fragebogen (beispielsweise *aggressives Verhalten beim Streicheln durch eine fremde Person*) und freundliches Verhalten im Wesenstest haben signifikant negativ korreliert mit Aggression im Fragebogen (beispielsweise *freundliches Verhalten beim freundlichen Annähern*). Auch für die anderen Messparameter gab es derartige Korrelationen, die Übereinstimmungen waren im Allgemeinen moderat.

Dennoch ist die Aussagekraft von Fragebögen im Sinne der Vorhersagbarkeit, beispielsweise zur Frühauswahl von Arbeitshunden, nicht ausreichend untersucht. Hier besteht deutlicher weiterer Forschungsbedarf.

#### **4.1.4 Conclusio Fragebogen**

Gegenüber Wesenstests haben Fragebögen jedenfalls logistische und finanzielle Vorteile (WIENER et HASKELL, 2016). Dennoch wird generell angenommen, dass sie weniger objektiv und präzise sind. BesitzerInnen fehlt im Vergleich zu trainierten BeobachterInnen meist das Verständnis für die große Palette des Hundeverhaltens, was zur angenommenen Voreingenommenheit hinzukommt. Fragebögen der/des VorbesitzerIn könnten auf der einen Seite verfälscht sein aufgrund von *“under-reporting”*, spiegeln aber eher das Verhalten in einem richtigen Zuhause wider als in einer Tierheimsituation (bei einem Wesenstest).

Die Übereinstimmung zwischen Vor- und NeubesitzerInnen und somit die Validität von Fragebögen setzt voraus, dass beide BesitzerInnen verlässliche (*reliable*) Angaben machen und dass der Hund sich in beiden Umgebungen einheitlich (*consistent*) verhält (STEPHEN et LEDGER, 2007). Zwei Zuhause (VorbesitzerIn vs. NeubesitzerIn) sind allerdings niemals identisch und die Vergleichbarkeit ist daher fraglich. Außerdem könnten die Antworten der VorbesitzerInnen teilweise von den Ansprüchen der NeubesitzerInnen abweichen.

Dennoch können die Ergebnisse einiger Messparameter hilfreich sein für das Tierheimpersonal und die Ausfülldauer des Fragebogens von ca. zwölf Minuten ist jedem zuzumuten. Fragebögen könnten daher eine Lösung sein, informatives Zusatzwissen über den Hund zu erlangen, allerdings sollten diese Informationen mit Vorbehalt gesehen werden, da noch immer die Frage der Subjektivität im Raum steht und das Verhalten eines Hundes nicht konsistent ist. Der Vorteil dürfte vor allem in jenen Aspekten liegen, in denen VorbesitzerInnen ein Problemverhalten auch tatsächlich angeben. Angaben des Nichtvorhandenseins eines bestimmten Problemverhaltens scheinen jedoch nur bedingt verwertbar zu sein.

Auch die Beeinflussung des Tierheimpersonals durch das Studieren der Abgabefragebögen wäre interessant, eventuell mit einer Kontrollgruppe, bestehend aus Fundtieren. Aber in der „Population von Fundhunden“ ist von einer anderen Vortestwahrscheinlichkeit auszugehen, da die Gründe von Abgabe und Aussetzen variieren könnten.

Untersuchungen des *C-BARQ* an Hunden in Privatbesitz sind vielversprechend, es taucht dadurch die Frage auf, was diese Hunde von der Tierheimhundepopulation unterscheidet, was vor allem die Intention der BesitzerInnen (*Abgabe im Tierheim vs. Teilnahme an einer Studie mit dem Familienmitglied*) und die Prävalenz der Verhaltensprobleme ausmachen dürfte. Somit stellt sich genau an dieser Stelle erneut die Frage nach der Objektivität der Beurteilung, welche die Voraussetzung für die Reliabilität und die Validität darstellt. Ob ein(e) nicht-abgebende(r) BesitzerIn ihren/seinen eigenen Hund völlig objektiv beurteilen kann, ist aber ebenfalls unsicher. Die Prävalenz für Verhaltensprobleme dürfte in den untersuchten Populationen stark variieren, in der Population initialen Studie zu potentiellen Assistenzhunden (SERPELL et HSU, 2001) ist sie wahrscheinlich niedriger als in der Population der Hunde, welche abgegeben werden. In der Folgestudie zu Hunden in Verhaltenstherapie (HSU et SERPELL, 2003) ist die Prävalenz für Verhaltensprobleme wahrscheinlich wiederum höher als in der Population der Hunde, welche in einem Tierheim abgegeben werden.

#### **4.2 Wesenstests**

Basierend auf den erhobenen Daten entspricht keiner der 13 untersuchten Wesenstests den Ansprüchen an ein diagnostisches Testverfahren.

Ein Vergleich der Studien bzw. Wesenstests ist aufgrund des Mangels an Standardisierung und Konsens sehr schwierig. In jeder Studie wurden andere Hunde inkludiert bzw. exkludiert.

Die Quelle der zu testenden Hunde war weitestgehend gleich – Hunde im Tierheim. Lediglich bezüglich des Geschlechts und Reproduktionsstatus wurden keine Exklusionsverfahren vorgenommen. Teilweise wurden aber Hunde bestimmter Rassetypen von vornherein ausgeschlossen oder Hunde mit potenziell aggressivem Verhalten wurden gar nicht erst getestet (z.B. in MORNEMENT et al., 2014).

Neben dem Ablauf, der unterschiedlichen Wahl der Messparameter, der Auswahl und Anzahl der BeurteilerInnen und der Auswahl der Hunde variierten die Studien auch im Ziel ihrer Forschung. In einigen Studien stand die Auswahl von Assistenzhunden im Vordergrund (z.B. WEISS et GREENBERG, 1997), in anderen Studien wiederum die Beurteilung der Vermittelbarkeit von Tierheimhunden (z.B. ZILOCCHI et al., 2015). Generalisierungen zwischen Arbeitshunden und Haushunden sollten jedoch mit Vorsicht vorgenommen werden (JONES et GOSLING, 2005). Außerdem könnten sich einige Messparameter in ihrer Relevanz bei der Auswahl von Assistenzhunden im Gegensatz zur Vergabe von Tierheimhunden unterscheiden.

Der Zeitpunkt des initialen Wesenstests und des Wiederholungstests, sofern dieser durchgeführt wurde, variierten ebenfalls stark zwischen den einzelnen Studien.

Bedenklich ist, dass der Wesenstest in Ländern ohne „no-kill-Politik“ (USA, Australien) meist innerhalb der ersten Tage stattfindet. Dies dürfte mit der Dringlichkeit einer zeitnahen Entscheidung über das weitere Vorgehen mit dem entsprechenden Hund zusammenhängen. Basierend auf den vorliegenden Daten, ist dies als Entscheidungsgrundlage für oder gegen eine Euthanasie jedoch äußerst kritisch zu sehen.

#### **4.2.1 Objektivität von Wesenstests**

Unter anderem differieren die Wesenstests in ihrem Ablauf. Grundsätzlich laufen die Wesenstests als sogenannte *Testbatterie* ab, das heißt als eine Kombination aus unterschiedlichen Einzeltests in einer meist definierten Reihenfolge. Dadurch soll die Objektivität gewährleistet werden. Einige der Wesenstests beginnen beispielsweise im Zwinger des Hundes (z.B. *Assess-A-Pet*) und in anderen Wesenstests wird der Hund auf das Testgelände geführt, wo der Wesenstest beginnt (z.B. Wesenstest der *RSPCA*). Die meisten Wesenstests finden drinnen statt, beispielsweise in einem Raum auf dem Tierheimgelände, den der zu testende Hund noch nicht kennt (CHRISTENSEN et al., 2007). In manchen

Wesenstests gibt es auch Einzeltests, die draußen in einem eingezäunten Bereich stattfinden oder während eines Spaziergangs (beispielsweise die Evaluierung der Leinenführigkeit bei *Match Up*, DOWLING-GUYER et al., 2011). Teilweise bestehen die Tests aus Kombinationen von Evaluierungen von drinnen und draußen (B.A.R.K., MORNEMENT et al., 2010). In vielen Studien wurde z.B. die Größe des Testgeländes genau definiert, diese variierte daher je nach Studie und auch die Genauigkeit der Beschreibung der Testumgebung ist unterschiedlich.

Auch die gewählten Messparameter unterschieden sich oft grundlegend und unterlagen starken Schwankungen hinsichtlich ihrer Objektivität. Beispielsweise kann die Messung von *Vokalisation* durch objektive Zählungen beurteilt werden, während *Gehorsam*, *Aktivitätslevel* oder *Aggression* einer eher subjektiven Einschätzung unterliegen (TAYLOR et MILLS, 2006). In einigen Aspekten der Verhaltensbeurteilung kann somit ein gewisses Maß an Standardisierung stattfinden (DIEDERICH et GIFFROY, 2006), in anderen scheint dies nicht der Fall zu sein.

Videoaufnahmen sind eine gute Ergänzung, um die Objektivität von Wesenstests zu erhöhen. Obwohl Videoaufnahmen als unterstützende Beweismittel bei manchen Wesenstests auch eingesetzt werden, kann die Beurteilung von Videoaufnahmen alleine nicht mit der Beurteilung vor Ort gleichgesetzt werden. Im Video ist immer nur ein Ausschnitt und nicht der gesamte Blickwinkel beurteilbar (PATRONEK et al., 2019), eine Steigerung der Objektivität allein durch Videoaufnahmen scheint daher nur bedingt möglich zu sein.

Zusätzlich muss man beachten, dass zwei Stimuli, welche auf uns Menschen gleich erscheinen, durch den Hund völlig unterschiedlich wahrgenommen werden können, da sich Hunde auf andere Eigenschaften fokussieren (TAYLOR et MILLS, 2006). So sind olfaktorische Reize für den Hund ein zentraler Punkt, wir sind uns diesen Reizen aber nur selten bewusst. Eine interessante Beobachtung haben POULSEN et al. (2010) gemacht: die Punktescores im Wesenstest der *RSPCA* waren relativ unabhängig vom Hundetyp und Umweltfaktoren, allerdings waren sie niedriger („besser“), wenn drei oder mehr Kothaufen von vorher getesteten Hunden auf dem Testareal vorhanden waren, für Harn waren Tendenzen in dieselbe Richtung vorhanden. Kot- und Harnabsatz durch den getesteten Hund selbst beeinflusste das Ergebnis dagegen nicht. Dies ist somit ein weiteres Beispiel, welches aufzeigt, dass die Standardisierung von Wesenstests relativ unmöglich erscheint.



Um die Praktikabilität zu steigern, könnten die Anzahl der Einzeltests reduziert werden und die Beobachtungsmethoden vereinfacht werden (TAYLOR et MILLS, 2006). Dadurch verbessert sich auch die Vereinheitlichung der Testabläufe und auch die Intra-Rater-, Inter-Rater- und Test-Retest-Reliabilität würden gesteigert werden. In der Realität werden allerdings ständig Abläufe, Beurteilungssysteme oder logistische Aspekte eines bereits beschriebenen Wesenstests modifiziert, beispielsweise werden Einzeltests weggelassen, modifiziert oder neue dazu genommen. Es gibt keine bis kaum eine Regulation oder professionelle Aufsicht darüber, ob solch ein „Missbrauch“ im Sinne einer Abwandlung einer beschriebenen Methodik stattfindet, auch wenn diese Modifikationen oft aus guter Absicht heraus geschehen. Die Validierung eines Wesenstests scheint daher unüberbrückbare Hindernisse mit sich zu bringen. Gäbe es einen validierten Wesenstest für die Tierheimpopulation, könnte dieser dann nicht beliebig modifiziert und den Gegebenheiten vor Ort angepasst werden, ohne dass jegliche auch minimale Änderungen neu validiert werden müssen (PATRONEK et al, 2019), eine scheinbar unlösbare Aufgabe in der Praxis eines Tierheims.

Die Betrachtung und der Vergleich der Testdauer der unterschiedlichen Wesenstests zeigten eine sehr hohe Varianz. Neunzig Minuten für einen Test (nach VAN DER BORG et al., 1991) aufzubringen, ist in einem Tierheimalltag eine sehr unrealistische Vorgabe. Das Wesen eines Hundes dagegen in nur drei Minuten zu erfassen (KIS et al., 2014), halte ich ebenfalls für nicht realistisch. Die Persönlichkeit und die Reaktionen von Lebewesen sind sehr komplex und abhängig von vielen Faktoren. Die Dauer des Tests der RSPCA war positiv korreliert mit dessen Bestehen (POULSEN et al., 2010). Dies war aber zu erwarten, da der Test bei starker Aggression oder extremer Angst abgebrochen wurde. Die größte Differenz in der Dauer gab es zwischen „Bestehen“ und „Verhaltensmodifikationsprogramm empfohlen“, vermutlich aufgrund der Verlängerung von Tests aufgrund von Unsicherheiten. Diese Erkenntnis spricht dafür, dass es schwer möglich ist, ein standardisiertes Verfahren mit einer festgelegten Dauer anzuwenden – und eine nicht festgelegte Dauer würde wiederum der Objektivität schaden.

In einer Befragung von Tierheimpersonal gaben 44% der Wesenstests durchführenden MitarbeiterInnen an, dass die vorgegebene Testdauer nicht ausreichend sei, um eine angemessene Beurteilung des Verhaltens eines jeden Hundes zu erhalten (MORNEMENT et

al., 2010). Daher wäre ein Ansatz, dass ein Teil der Informationen vielmehr aus dem täglichen Umgang erfasst wird und bei Bedarf einige zusätzliche Tests durchgeführt werden, welche optional bzw. auch auf den Bedürfnissen des/der potentiellen NeubesitzerIn basierend ergänzt werden.

Sowohl die Anzahl, als auch der Ausbildungsstand der testenden Personen differierte in den verschiedenen Wesenstests dramatisch. Individuelle Unterschiede im Handling der Personen könnten das Testergebnis unabsichtlich verändern (TAYLOR et MILLS, 2006). In einer Befragung in Australien gaben 59% der Tierheimangestellten (welche auch Wesenstests durchführen) an, ihr Wissen einzig im Berufsalltag erlangt zu haben und nur 33% haben an Seminaren oder Kursen zu Hundeverhalten teilgenommen (MORNEMENT et al., 2010). In der Praxis dürfte das fundierte Wissen der PrüferInnen demnach starken Schwankungen unterliegen und von der individuellen Eigeninitiative und der persönlichen Einstellung der MitarbeiterInnen abhängen. Auch zu angewandten Trainingsmethoden gibt es keine einheitlichen Vorgaben. Der größte Teil der Befragten (50%) gab an, sich bei der Beurteilung von Hunden im Wesenstest „einigermaßen kompetent“ zu fühlen. Interessanterweise war die Erfahrung der PrüferInnen nicht korreliert mit deren Selbstvertrauen in ihre Fähigkeit, Hunde exakt beurteilen zu können. Das Personal ist mehr oder weniger gezwungen, Entscheidungen über Leben und Tod zu treffen auf Grundlage von möglicherweise inadäquaten Trainingsmethoden, potentiell ungültigen Wesenstests und subjektiven Interpretationen über Verhalten, mit denen sich die Personen nicht in einem hohen Maße wohl fühlen. Auch die Auswirkungen auf die Psyche der MitarbeiterInnen sollten dabei nicht unerwähnt bleiben. Eine Verbesserung dieser Situation könnte durch kontinuierliche Aus- und Weiterbildung von Tierheimpersonal als auch durch eine Vereinheitlichung von Handlingkriterien erfolgen.

#### **4.2.2 Reliabilität von Wesenstests**

Es ist erschreckend, wie unzureichend die Reliabilität von Wesenstests untersucht wurde, trotz dessen, dass diese Evaluierungsmethode bereits seit Jahrzehnten in vielen Tierheimen weltweit Anwendung findet. Die Reliabilität ist Voraussetzung für die Validität von Testverfahren und somit existenziell.

Die Tatsache, dass in manchen Studien Hunde mit Aggressionsvorgeschichte von vornherein exkludiert wurden oder nur jene Hunde einbezogen wurden, die geeignet für die Vermittlung

erschienen, steigert die Homogenität des Verhaltens der getesteten Hunde. Durch eine verminderte Varietät wird das Erreichen eines gewissen Maßes an Reliabilität erschwert. Umgekehrt, werden Hunde mit Aggressionsvorgeschichte oder offensichtlichem Drohverhalten inkludiert, könnte dies die Reliabilität fälschlicherweise erhöhen, da die typische Vergabepopulation an Hunden in Ländern ohne *no-kill-Politik* diese Hunde normalerweise nicht beinhaltet und die Reliabilität aber für genau diese zu vermittelnde Hundepopulation gefragt ist (PATRONEK et al., 2019). Länderspezifisch unterschiedliche Legislationen verkomplizieren diese Problematik daher noch zusätzlich.

#### 4.2.2.1 Test-Retest-Reliabilität von Wesenstests

Nur sechs AutorInnen haben den Test wiederholt (siehe Tab. 29), dabei hat die Test-Retest-Reliabilität stark variiert je nach Wesenstest, Messparameter, Zeitabstand zwischen den Tests und Gegebenheiten im Wiederholungstest. Der Abstand zwischen Test und Retest lag dabei zwischen 24 Stunden (MORNEMENT et al., 2014) und über 80 Tagen (POULSEN et al., 2010). Ein direkter Vergleich ist daher kaum möglich. Nur wenige Einzeltests bzw. Messparameter hatten eine hohe Korrelation, beispielsweise ängstliche Verhaltensweisen (KIS et al., 2014; MORNEMENT et al., 2014). Interessant wären mehrere Wiederholungstests innerhalb des Tierheimaufenthaltes. Die Evaluierung der Test-Retest-Reliabilität von Hunden im Tierheim birgt allerdings auch Probleme in der Logistik und in der Sinnhaftigkeit (PATRONEK et al., 2019).

Einige der AutorInnen haben den Retest analog zum ersten Test im Tierheim abgehalten, andere AutorInnen haben den Wiederholungstest erst bzw. zusätzlich nach der Adoption durchgeführt (VALSECCHI et al., 2011; POULSEN et al., 2020).

Hierbei haben VALSECCHI et al. (2011) die Wiederholung des *Temperament Tests* sowohl im Tierheim, als auch im neuen Zuhause durchgeführt – die Ergebnisse (deutliche vs. moderate Übereinstimmung) sprechen dafür, dass ein Retest im neuen Zuhause eine weit weniger starke Übereinstimmung erreicht. Der Punktescore im neuen Zuhause ist signifikant gestiegen und hat sich somit verbessert. Auch durch POULSEN et al. (2010) wurde nach der Adoption im Wiederholungstest (Wesenstest der *RSPCA*) in vielen Einzeltests keine Übereinstimmung erreicht. Diese Ergebnisse zeigen auf der einen Seite, dass ein Wechsel der Umgebung (Tierheim vs. neues Zuhause) die Wiederholpräzision vermindert, auf der

anderen Seite werfen sie aber auch erneut die Frage auf, wie sinnvoll ein Wesenstest ist, um Reaktionen zu messen, welche stark von den örtlichen Gegebenheiten abzuhängen scheinen.

Auch die Ergebnisse der Wiederholungstests im Tierheim selbst variieren stark bezüglich der Messparameter. *Angst* scheint sowohl nach 24 Stunden (MORNEMENT et al., 2014; B.A.R.K.), als auch nach zwei Wochen (KIS et al., 2014; Wesenstest nach KLAUSZ et al.) als einziger gemessener Parameter eine hohe Übereinstimmung zu erreichen. Andere Messparameter scheinen trotz kurzer Abstände zwischen den Tests nur eine niedrige Übereinstimmung zu erreichen, beispielsweise *Hundeverträglichkeit* (BENNETT et al., 2015; SAFER™) erlangt nach nur drei Tagen eine lediglich ausreichende Übereinstimmung. Sogar *Freundlichkeit* erreicht nach nur 24 Stunden (MORNEMENT et al., 2014; B.A.R.K.) eine nur moderate Übereinstimmung. Auch die Punktescores in der Studie von POULSEN et al. (2010; Wesenstest der RSPCA) waren im Wiederholungstest insgesamt niedriger („besser“), der Unterschied war allerdings nicht signifikant. Dies könnte dafür sprechen, dass die Hunde erst nach einer Akklimatisierungsphase im Tierheim ihr eigentliches Verhalten zeigen. Ein Testzeitpunkt innerhalb dieser Akklimatisierungsphase dürfte also nur bedingt sinnvoll und aussagekräftig sein. Eine Möglichkeit wäre, den Wesenstest z.B. im Monatstakt zu wiederholen, um die Entwicklung des Verhaltens über die gesamte Aufenthaltsdauer im Tierheim zu beobachten und auszuwerten, doch würde dies wiederum schwer praktikabel sein.

Die Schwierigkeiten liegen hier im Auslösen ähnlicher Verhaltensweisen beim erneuten Testen, in der Unmöglichkeit der wissenschaftlichen Bestimmung des idealen Wiederholungstestzeitpunkts und in der Unsicherheit über die Auswirkung von Anpassung oder Nichtanpassung des Hundes an das Tierheim und das Tierheimpersonal. Auch die Frage des Lerneffektes des Hundes von einem auf den nächsten Wesenstest bleibt offen. Ein Wiederholungstest wird daher nicht viel mehr aussagen, als des Hundes spezifische Reaktion zu einem spezifischen Zeitpunkt an einem spezifischen Ort (PATRONEK et al., 2019).

Ein möglicher Grund für die starken Diskrepanzen zwischen den Tests ist eine Verhaltensentwicklung während des Aufenthaltes im Tierheim. Im Folgenden werde ich daher auf diesen Punkt näher eingehen. MORNEMENT et al. (2014) haben leider keine Verhaltensentwicklung vom ersten zum zweiten Test nach B.A.R.K. erfasst.

In der Studie von BENNETT et al. (2015) zum SAFER™-Test hatten die Hunde nur bei den Einzeltests „Berührungssensitivität“, „Spielverhalten“ und im „Kneiftest“ im Wesenstest nach drei Tagen niedrigere Level an Aggression. Ob dies einem niedrigeren Stresslevel nach drei Tagen oder einem Lerneffekt vom ersten Test zuzuschreiben ist, ist unklar. In den anderen Einzeltests konnte keine Richtung der Verhaltensentwicklung festgestellt werden.

Die Hunde zeigten im Wesenstest nach KLAUSZ et al. nach zwei Wochen mehr Aggression als nach ein bis zwei Tagen (KIS et al., 2014). Dies war allerdings nur in jenem Einzeltest der Fall, in dem Futter weggenommen wurde. Als mögliche Gründe wurden Angst und supprimierte Verhaltensreaktionen durch den erhöhten Kortisolspiegel in den ersten Tagen genannt, sowie das initiale Vorziehen von Meideverhalten statt Konfrontation aufgrund der Umgebungsveränderung. Der Wesenstest von KIS et al. (2014) unterscheidet sich von anderen untersuchten Wesenstests allerdings gravierend, da der Hund an zwei Ketten so angebunden ist, dass er lediglich etwas Spielraum nach vorne und hinten hat. Jede Interaktion findet ausschließlich mit der *Fake-Hand* statt. Diese Vorgehensweise könnte für den zu testenden Hund sehr einschüchternd wirken bzw. seine Reaktion auf Einengung stark provozieren.

FRANK berichtet bei der Studie zu ihrem Wesenstest von einer positiven Verhaltensänderung, da im Gegensatz zu ca. 27% der Hunde im Eingangstest nur 17.4% der Hunde im Wiederholungstest nach vier Wochen auffälliges Verhalten zeigten. Allein der Einzeltest „auf den Rücken drehen“ wies im Wiederholungstest einen Anstieg von aggressivem Verhalten gegenüber der Testerin auf.

Diesen recht spärlichen Informationen könnte man entnehmen, dass die Toleranzschwelle für provokative Einzeltests (*Futter wegnehmen, auf den Rücken drehen*) prinzipiell sinkt und in diesen Tests die Bereitschaft für Aggression steigt. Die Einsatzbereitschaft von *Aggression* als Lösungsstrategie in weniger provokativen Einzeltests (*Berührung, Spielen*) scheint dagegen zu sinken.

Laut POULSEN et al. (2010) beeinflusste die Dauer des bisherigen Tierheimaufenthaltes die Ergebnisse des Wesenstests (der RSPCA) nicht. Der Gesamtpunktescore der beiden Tests 20 und 60 Tage nach Aufnahme im Tierheim (VALSECCHI et al., 2011) ist nicht signifikant gestiegen vom ersten zum zweiten *Temperament Test*.

Keine der Studien hat die Auswirkungen von deutlich höheren Zeitspannen zwischen den beiden Tests untersucht, der längste Abstand lag bei 40 Tagen (VALSECCHI et al., 2011) zwischen Tests während des Tierheimaufenthaltes. Wie Hunde sich in ihrem Verhalten entwickeln, nach beispielsweise einem oder mehreren Jahren im Tierheim, wurde in den Studien nicht untersucht. Auch wie stark die Korrelationen für z.B. den Messparameter *Angst* bei einem Wiederholungstest nach einem Jahr noch sind, wurde bis dato von niemandem betrachtet.

Nur wenn ein Verhalten in wiederholten Tests konsistent ist, kann dieses Verhalten auf das Temperament des Hundes zurückgeführt werden. Hunger, Krankheit und andere Gründe können das Verhalten jedoch von Test zu Test verändern. Auch die Reaktion auf einen neuen Stimulus lässt sich schwierig wiederholen. Im Allgemeinen ist ein Wesenstest dem Hund neu, während ihm alle Elemente (PrüferIn, Ort, Objekte) bei einer Wiederholung bereits vertrauter sind, was das Verhalten beeinflussen könnte. Durch Gewöhnung könnte ein Hund weniger aggressiv reagieren oder auch aggressiver durch den „Lernerfolg“ im initialen Test. Die Art und Weise, was der Hund aus dem ersten Test gelernt hat, kann natürlich auch etwas über dessen Wesen aussagen. Je weniger Zeit zwischen den beiden Tests liegt, desto weniger könnten die Stimuli als „neue Reize“ wirken, welche sie in einer Wiederholung ohnehin nicht mehr sein können (TAYLOR et MILLS, 2006).

Auch hier wollen wir die Literatur bei Hunden in Privatbesitz betrachten. KLAUSZ et al. (2013) haben die Test-Retest-Reliabilität des Wesenstests *nach KLAUSZ et al.* bei Hunden in Privatbesitz untersucht, es wurden keine signifikanten Unterschiede gefunden zwischen den Tests. VAN DER BORG et al. (2010) haben die Test-Retest-Reliabilität des *SAB*-Tests untersucht. Allerdings wurden die beiden Tests an unterschiedlichen Orten durchgeführt (Innen- vs. Außenbereich). Die Test-Retest-Reliabilität hat für *Aggression* mit 0.55 signifikant korreliert und für *Angst* ebenfalls mit 0.82. Der Ort schien keinen signifikanten Einfluss zu nehmen, die Punktezahlen lagen für *Aggression* und *Ängstlichkeit* im ersten Test am Morgen aber höher als im zweiten Test am Nachmittag. SVARTBERG et al. (2004) haben für den *DMA* (Dog Mentality Assessment) die Test-Retest-Reliabilität untersucht, indem der Wesenstest dreimal analog durchgeführt wurde, jeweils im Abstand von etwa 30 bis 35 Tagen zwischen den Tests. Die/der BesitzerIn hat den Hund während des Wesenstests geführt. Die teilnehmenden Hunde waren ebenfalls Hunde in Privatbesitz. Die Korrelationen

variierten je nach Messparameter und Wiederholung, waren aber alle signifikant. Im Messparameter *Mut* variieren die Übereinstimmungen zwischen 0.83 (1. vs. 3. Test) und 0.90 (2. vs. 3. Test), im Messparameter *Verspieltheit* zwischen 0.76 (1. vs. 3. Test) und 0.89 (2. vs. 3. Test), im Messparameter *Jagdtrieb* zwischen 0.61 (1. vs. 3. Test) und 0.80 (2. vs. 3. Test), im Messparameter *Neugier* zwischen 0.58 (1. vs. 3. Test) und 0.75 (2. vs. 3. Test), im Messparameter *Umgänglichkeit* zwischen 0.57 (1./2. vs. 3. Test) und 0.72 (1. vs. 2. Test) und im Messparameter *Aggressivität* zwischen 0.68 (1. vs. 2./3. Test) und 0.80 (2. vs. 3. Test). Es sind somit klare Tendenzen erkennbar, dass die Testwiederholungen mit der Anzahl stärker korrelieren, da die Übereinstimmungen zwischen dem 2. und 3. Test am höchsten sind. Dies wäre ein weiterer Hinweis auf stattfindende Verhaltensänderungen bedingt durch den Lerneffekt von Hunden, in diesem Fall über einen Zeitraum von drei Monaten.

In einer anderen Studie wurde die Test-Retest-Reliabilität bei Laborhunden (*Beagles*) getestet (DÖRING et al., 2017a). Der Wiederholungstest fand bereits einen Tag nach dem initialen Test statt. Die Korrelationen waren niedrig bis moderat. Interessant ist, dass die Verhaltensreaktionen und die Körpersprache separat beurteilt wurden. Die Körpersprache erzielte in den meisten Beurteilungspunkten weit stärkere Übereinstimmungen und kann daher als konsistenter innerhalb eines Individuums betrachtet werden, als die Reaktionen selbst, welche stark von der jeweiligen Situation und dem jeweiligen Stimulus abhängig waren. Somit schließt sich auch hier der Kreis - wenn von den beurteilenden Personen eine regelmäßige Aus- und Weiterbildung gefordert wird, sind diese in der Lage, auf die Körpersprache selbst Rücksicht zu nehmen, anstatt nur Reaktionen von Hunden zu messen. In der Folgestudie (DÖRING et al., 2017b) wurde ein Teil der Hunde sechs Wochen nach der Vermittlung im neuen Zuhause ein weiteres Mal getestet. Zwischen den beiden Testungen wurden keine signifikanten Unterschiede bezüglich der Verhaltensreaktionen gemessen. Die Beurteilung der Körpersprache der Hunde änderte sich allerdings signifikant – im neuen Zuhause erlangte sie einen entspannteren Zustand.

#### **4.2.2.2 Inter-Rater-Reliabilität von Wesenstests**

Sechs der Studien haben die Inter-Rater-Reliabilität untersucht. Dabei ist festzustellen, dass die Inter-Rater-Reliabilität von der Anzahl der BeurteilerInnen und den untersuchten Messparametern abhängt.

In fünf von sechs Studien wurde die Übereinstimmung zwischen zwei BeurteilerInnen untersucht, die Übereinstimmung war dabei mindestens ausreichend (MATTERS, 2016, Wesenstest der ASPCA®; VALSECCHI et al., 2014, *Temperament Test*) und reichte sogar bis hin zu (nahezu) perfekter Übereinstimmung (KIS et al., 2014, Wesenstest nach KLAUSZ et al.; VALSECCHI et al., 2011; MATTERS, 2016). Nur DIESEL et al. (2008, *Dogs Trust*) haben die Übereinstimmung zwischen 40 BeurteilerInnen untersucht und dabei schwache bis maximal moderate Übereinstimmung feststellen können.

Zwischen routinierten BeurteilerInnen (*current raters* nach MATTERS, 2016, Wesenstest der ASPCA®) ist die Korrelation stärker als zwischen nicht-routinierten BeurteilerInnen (*non-current raters*). Zwischen sehr erfahrenen BeurteilerInnen ist die Übereinstimmung ebenfalls größer als zwischen weniger erfahrenen BeurteilerInnen (DIESEL et al., 2008, *Dogs Trust*). Diesen Informationen ist zu entnehmen, dass für die Beurteilung von Wesenstests sowohl sehr erfahrene BeurteilerInnen notwendig sind, diese aber auch über angewandte Routine in der Beurteilung verfügen sollten.

PATRONEK et al. (2019) erwähnen jedoch Abstufungen der Übereinstimmung je nach Komplexität des getesteten Verhaltens. Während einfache Verhaltensweisen (z.B. *Verspielt-heit*) bei VALESECCHI et al. (2011, *Temperament Test*) eine starke Übereinstimmung erreichen, ist die Übereinstimmung bei komplexeren Verhaltensweisen (z.B. *Umgänglichkeit*) wesentlich schwächer.

Leider haben einige AutorInnen die Übereinstimmung bezüglich bestimmter Messparameter (z.B. *Angst*) und andere AutorInnen bezüglich Einzeltests (z.B. *Reaktion auf ein Spielzeug drinnen* bei MATTERS, 2016, Wesenstest der ASPCA®) gemessen, was einen direkten Vergleich der Wesenstests erschwert. DIESEL et al. (2008, *Dogs Trust*) haben die Messparameter nach Situation aufgeschlüsselt (z.B. *angst aggressives Verhalten bei Annähern an den Zwinger*). Würde man die Reliabilität nach Situation aufschlüsseln, würde die Testsituation nur dann Sinn machen, wenn die Reliabilität für die Beurteilung jeder möglichen Reaktion hoch wäre. Aber auch hier herrscht starke Varianz: beispielsweise reichen die Korrelationen bei der Beurteilung des Verhaltens bei Annähern an den Zwinger von schwach (*gleichgültige Reaktion*) bis erheblich (*nervöse Reaktion*). Daraus kann man schließen, dass es verlässlich ist, wenn eine *nervöse Reaktion beim Annähern an den Zwinger* beobachtet wird, wird dagegen aber eine *gleichgültige Reaktion bei Annähern an*



den *Zwinger* erfasst, ist diese Beobachtung stark von der/dem BeurteilerIn abhängig und somit kaum zuverlässig. Ähnliches zeigte sich bei der Reaktion beim Treffen eines Artgenossen. Lediglich schwache (*offensiv-aggressive Reaktion*) bis ausreichende Übereinstimmungen wurden gefunden, nur die *aufgeregte Reaktion* erreichte eine moderate Übereinstimmung (DIESEL et al., 2008, *Dogs Trust*).

Die Beurteilung des Verhaltens des Hundes alleine im Außenzwinger variierte ebenfalls zwischen keiner und moderater Übereinstimmung (MATTERS, 2016, Wesenstest der ASPCA®). Da in dieser Studie vier BeurteilerInnen in allen Kombinationen verglichen wurden, ist hier besonders die starke Varianz abhängig vom BeurteilerInnenpaar sichtbar.

Interessanterweise konnte eine sehr hohe Übereinstimmung beim Erkennen von *Angst* festgestellt werden (MORNEMENT et al., 2014; *B.A.R.K.*). KIS et al. (2014, Wesenstest *nach KLAUSZ et al.*) haben *Angst* beurteilt und dabei nahezu eine perfekte Übereinstimmung erreicht. Dabei ist allerdings zu erwähnen, dass sich dieser Wesenstest (wie bereits beschrieben) im Ablauf stark von allen anderen unterscheidet.

Ein weiteres Beispiel für die starke Varianz der Messergebnisse ist *Fügsamkeit an der Leine*, welche durch VALSECCHI et al. (2011; *Temperament Test*) als nicht signifikant übereinstimmend beurteilt wurde. *Leinenführigkeit* erreichte allerdings bei MATTERS (2016, Wesenstest der ASPCA®) abhängig von der Kombination der BeurteilerInnen moderate bis (nahezu) perfekte Übereinstimmung.

Es ist daher nicht verwunderlich, dass PATRONEK et al. (2019) die Notwendigkeit einer Messung einer sogenannten „Inter-Tierheim-Reliabilität“ aufgrund der starken Heterogenität diverser Tierheimumgebungen erwähnen.

Auch hier wird erneut auf die Sinnhaftigkeit von Videoaufnahmen zur Erfassung der Inter-Rater-Reliabilität hingewiesen, da eine hohe Anzahl an BeurteilerInnen die/den PrüferIn und den Hund beeinflussen können.

VAN DER BORG et al. (2010) haben die Inter-Rater-Reliabilität des *SAB*-Tests bei Hunden in Privatbesitz untersucht. Die Inter-Rater-Reliabilität wurde zwischen drei BeurteilerInnen mit jeweils über zehn Jahren Erfahrung im Beurteilen von *SAB tests* gemessen. Für Aggression waren die Übereinstimmungen hoch und signifikant zwischen 0.86 und 0.87, für Angst waren

sie moderat und signifikant zwischen 0.66 und 0.67. DÖRING et al. (2016) haben einen standardisierten zehnteiligen Wesenstest für Laborbeagles entwickelt und dabei eine hohe Inter-Rater-Reliabilität von 93.1% erlangt. Eine hohe Inter-Rater-Reliabilität könnte für das jeweilige Testverfahren und dessen Auswahl der BeurteilerInnen sprechen, da dieser Parameter unabhängig von der Tierheimsituation sein könnte – die An-/Abwesenheit der BesitzerInnen könnte aber auch hier eine wesentliche Rolle spielen.

#### **4.2.2.3 Intra-Rater-Reliabilität von Wesenstests**

In nur drei Studien wurde die Intra-Rater-Reliabilität untersucht. Erwartungsgemäß war die Intra-Rater-Reliabilität allgemein stärker als die Inter-Rater-Reliabilität. Dennoch erreichten beispielsweise die *nervöse* und die *brave Reaktion* beim *Treffen von Artgenossen* nur eine schwache Übereinstimmung (DIESEL et al., 2008, *Dogs Trust*). Auch die anderen möglichen Reaktionen in dieser Testsituation erreichten maximal eine moderate Übereinstimmung. In den anderen Situationen (*Annähern an den Zwinger, allgemeines Handling*) gab es Reaktionen, welche nur ausreichend übereinstimmten zu verschiedenen Zeitpunkten. Dieses Erkenntnis halte ich für erschreckend, da es sich um die Beurteilung durch ein und dieselbe Person handelt. In den anderen beiden Studien waren die Ergebnisse vielversprechender, wobei bei VALSECCHI et al. (2011, *Temperament Test*) festzustellen war, dass die Übereinstimmung mit der Zeitspanne zwischen den Beurteilungen (1 Monat vs. 16 Monate) abnahm.

#### **4.2.3 Prognostische Validität von Wesenstests**

Die Voraussage von Problemverhalten nach der Adoption ist wesentliche Indikation für die Durchführung von Wesenstests (CLAY et al., 2020c). Die Follow-Up-Studien sind eine Voraussetzung für eine Messung dieser prognostischen Validität. Ein Vergleich der Studienergebnisse ist aber auch hierbei nicht ausreichend möglich, da die Follow-Up-Studien nicht zum gleichen Zeitpunkt stattfanden. Die Zeitspanne variierte von wenigen Tagen bis zu über einem Jahr (siehe Tab. 5). Eine kurze Zeit zwischen Vergabe und Interview der NeubesitzerInnen könnte den negativen prädiktiven Wert (da manches Verhalten vielleicht noch nicht gezeigt wurde) und die Zahl der Falsch-Positiven verfälschen (VAN DER BORG et al., 1991). Außerdem wurden manche Situationen durch die NeubesitzerInnen nie provoziert und die An-/Abwesenheit des Verhaltens ist daher nicht messbar, was wiederum die Sensitivität des Tests beeinflusst.

Leider beträgt auch die Teilnehmerate an den Follow-Up-Studien in den vorhandenen Untersuchungen nur einen Bruchteil der im Tierheim getesteten Hunde, siehe Tab. 6. Die Gründe können vielseitig sein und inkludieren, dass die NeubesitzerInnen Probleme mit dem Hund haben, dies aber nicht eingestehen wollen, zu beschäftigt sind oder den Hund bereits an einen anderen Platz weitergegeben haben (TAYLOR et MILLS, 2006). Das Vorhandensein von Follow-Up-Studien mit einer Teilnehmerate von über 80% ist jedoch eine der Voraussetzungen für die Einstufung einer Studie in die Evidenzklasse Platin. Die Evidenz aller eingeschlossenen Untersuchungen kann somit als unzureichend bezeichnet werden. Durch das kontrollierte Durchführen von Follow-Up-Studien könnte die Studienqualität und damit deren Evidenz und Aussagekraft verbessert werden. Ebenso könnte eine mehrfache Durchführung über einen längeren Erfassungszeitraum helfen, Verhaltensentwicklungen und Verhaltensveränderungen, vor allem von einzelnen Individuen, besser zu erfassen. Nicht zu vergessen ist, dass es bis dato keinen Goldstandard zur Messung von Verhalten gibt. Auch klinisch haben sich keine Referenztests etabliert, beispielsweise erlauben Messungen von Kortisol- oder anderen Blutwerten keine eindeutigen Rückschlüsse auf Verhaltensprobleme (PATRONEK et al., 2019). Daher ist in der Tierheimsituation nur das Verhalten nach der Vergabe als Referenzwert sinnvoll. Weiters könnte ein Wesenstest vor der Vergabe wiederholt werden, bzw. in zeitlich definierten Abständen im Tierheim. Ob dies aber als Referenzwert zur Vorhersage von Verhalten im neuen Zuhause dienen kann, ist dennoch fraglich.

Ein wichtiger Parameter bei der Berechnung der prognostischen Validität ist die Prävalenz. Die Konsequenzen einer Vernachlässigung der Prävalenz wirken sich beispielsweise auf die Vorhersagekraft eines positiven Tests aus (PATRONEK et BRADLEY, 2016). Ein Beispiel hierfür wäre das Auslösen von Beißen oder Warnverhalten des Hundes im Wesenstest. Ein grundlegendes Problem ist hierbei, dass Bissstatistiken sowohl unzuverlässig sind (aufgrund des Mangels an Berichterstattung), als auch stark variieren nach Definition des Bisses und daher Schätzungen zur Prävalenz von beispielsweise erfolgten Bissen durch Hunde sehr vage sind. Ob die Prävalenz der Gesamtpopulation auf die Population der Tierheimhunde übertragbar ist, ist ebenfalls fraglich, da ein Teil der Hunde im Tierheim aufgrund von Verhaltensproblemen abgegeben oder ausgesetzt wurde und die entsprechende Prävalenz daher höher sein könnte. PATRONEK und BRADLEY (2016) haben Prävalenzangaben von

*Aggression* bei Tierheimhunden in vier unterschiedlichen Studien zwischen 11.3% und 16% gefunden.

Nachdem Hunde teilweise bereits bei Aufnahme im Tierheim durch Wesenstests gescreent werden, wird die gesamte Tierheimhundepopulation einem provokativen Test ausgesetzt, welcher auf Annahmen nicht existierender Daten zur Beurteilung von Hundeverhalten aufbaut und dadurch nicht haltbar ist. Es gibt außerdem keinerlei wissenschaftliche Informationen darüber, welches Level an Provokation zum Auslösen von Beiß- und Warnverhalten vom Hund tatsächlich als *normal* bezeichnet werden kann (PATRONEK et al., 2019). Weiters wurden die meisten Provokationstests, ihre Rechtfertigung und ihre anschließenden Interpretationen empirisch nicht evaluiert. Einige von ihnen werden trotz bereits bestehender Zweifel an ihrer Tauglichkeit als Testmodell in einigen Wesenstests regulär verwendet (beispielsweise die Kleinkindpuppe).

Das Ziel eines klinischen diagnostischen Tests ist es, festzustellen, ob ein Subjekt über einen bestimmte Eigenschaft verfügt oder nicht (PATRONEK et BRADLEY, 2016). In einem Wesenstest würde die „Diagnose“ nicht nur beinhalten, ob der Hund das Verhalten in einem oder mehreren Tests im Tierheim zeigt, sondern auch, dass das Verhalten, falls es gezeigt wurde, eine feste Eigenschaft ausmacht, die auch in anderen Kontexten auftritt und eine Gefahr darstellt. Im unwahrscheinlichen Fall, dass der erste Fakt eine hohe Reliabilität erreichen würde, bleiben die zwei anderen Fakten vollkommen spekulativ.

DOWLING-GUYER et al. (2011) definieren die Persönlichkeit als Kombination aus genetischen, kognitiven und äußeren Faktoren und sehen die Persönlichkeit als zumindest teilweise Grundlage für Verhaltenstendenzen in verschiedenen Situationen. Erfasst man die Persönlichkeit eines Hundes, sollen sich so Reaktionstendenzen auf Stimuli vorhersagen lassen. Diese Hypothese ist jedoch in der Studie nicht bewiesen worden, da keine Referenzwerte im neuen Zuhause gemessen wurden.

Testbatterien bergen im Vergleich zu Einzeltests ihre eigenen Spannungsfelder hinsichtlich Sensitivität und Spezifität (PATRONEK et BRADLEY, 2016). Für den Fall, dass ein positiver Einzeltest reicht, um in der Testbatterie insgesamt als positiv gewertet zu werden, ist die Gesamtsensitivität einer Testbatterie größer als für den einzelnen Test, aber die Gesamtspezifität ist geringer und somit die Zahl der Falsch-Positiven. Für den Fall, dass mehrere

Einzeltests positiv sein müssen, um in der Testbatterie positiv evaluiert zu werden, ergibt sich daraus: Die Gesamtspezifität ist größer als für den einzelnen Test, aber die Sensitivität ist geringer. Es macht weiters einen Unterschied, ob parallel getestet wird (der Folgetest hängt nicht vom vorherigen Test ab) oder in einer Serie von Tests. Es ist daher nicht überraschend, dass PATRONEK und BRADLEY (2016) unter Berücksichtigung der Kombination aus Sensitivität, Spezifität und Prävalenz von Beißen und Warnverhalten in einem positiven Wesenstestergebnis keine genauere Vorhersagekraft als jene durch einen Münzwurf sehen, wenn sie aufgrund der hohen Zahl an falsch-positiv identifizierten Hunden nicht sogar noch schlechter ist als 50/50.

Ein weiteres Problem stellt die Tatsache dar, dass einige AutorInnen einen großen Umfang an Einzeltests und Messparametern versuchen abzudecken, was die Sensitivität des gesamten Wesenstests vermindern könnte (TAYLOR et MILLS, 2006).

#### **4.2.3.1 Prognostische Validität ausgewählter Messparameter**

##### **4.2.3.1.1 Aggression gegenüber Menschen**

Das Grundmuster von aggressivem Verhalten scheint unveränderlich im genetischen Code von Hunden verankert zu sein (COREN, 2004). Zusätzlich zur genetischen Komponente wird der individuell gezeigte Grad an Aggression durch frühe Erfahrungen (*Sozialisation*) und Gelerntes geformt. Jeder Hund kann aggressives Verhalten zeigen, es ist Teil des Ausdrucksverhaltens aller Hunde, die Neigung dies auch zu tun, kann aber erheblich variieren. Hunde zeigen Aggression unter anderem in Konflikt- und Wettbewerbssituationen mit anderen Tieren, inklusive dem Menschen. Aggression reicht von milden Formen (*Erstarren, Knurren*) bis hin zu schweren Formen (*Beißen, Herausreißen von Fleisch, Tod*). Trotz der wichtigen kommunikativen Funktion, kann Aggression für alle Beteiligten sehr gefährlich werden und genießt kaum Akzeptanz in unserer Gesellschaft. Besonders unangemessene Reaktionen des Menschen auf Zeichen von Aggression beim Hund können zu einer folgenschweren Eskalation führen (BOLLEN et HOROWITZ, 2008).

An dieser Stelle ist die mangelnde Konsistenz des Begriffs „*Aggression*“ zwischen den Studien zu erwähnen. Aggression ist eine heterogene Klasse von Körperhaltungen und Aktionen, welche ein Teil des normalen Verhaltensrepertoires des Hundes sind und in einem großen Spektrum an Frequenz und Intensität über die Zeit hinweg auftauchen können, in

verschiedenen Gegebenheiten und mit verschiedenen Stimuli (PATRONEK et BRADLEY, 2016). Auch scheinbar konkretere Beschreibungen, wie beispielsweise „Knurren“ oder „Beißen“ sind situationsbedingt völlig unterschiedlich zu bewerten, ob es sich zum Beispiel um einen spielenden Welpen handelt oder einen erwachsenen Hund in einer anderen Situation. Welche Körperhaltungen oder Handlungen in den Bereich „*Aggression*“ fallen, differiert zwischen den Studien, ebenso wie im individuellen Verständnis von Personen. Beispielsweise bedeutet ein Knurren des Hundes als Reaktion auf eine fremde Person, welche den Testraum betritt, weder automatisch, dass das Verhalten des Hundes abnormal ist, noch dass ein/e neue/r BesitzerIn dieses Verhalten unter diesen Umständen als problematisch sehen würde, noch dass dieses Verhalten ein Problem im neuen Zuhause darstellen wird. Der Grenzwert dafür, was für eine Vermittlung noch verantwortet werden kann, variiert ebenfalls von einem einzelnen Knurren in einem Einzeltest der Testbatterie bis hin zu mehreren Bissversuchen der *Fake-Hand* und unterliegt somit eher der jeweiligen Tierheimpolitik und den zugrundeliegenden Umständen als der Wissenschaft.

Hinzu kommen noch individuelle Meinungen: zwischen den Hunderassen gibt es Unterschiede hinsichtlich dessen, was als „normal“ oder „akzeptabel“ bezeichnet werden kann (TAYLOR et MILLS, 2006), während ursächliche Gründe wie Furcht oder Frustration bzw. allgemeine Verunsicherung durch die Tierheimsituation meist vollkommen unbeachtet bleiben. Inwiefern dies in Wesenstests berücksichtigt wird/ werden kann, ist jedoch noch unklar.

In einer Befragung von Tierheimpersonal wurden folgende Verhaltensprobleme als die drei relevantesten Probleme bei Hunden im Tierheim aufgelistet: innerartliche Aggression, Angst und Ungehorsam (MORNEMENT et al., 2010). Aggression gegenüber Menschen tauchte in dieser Befragung nicht unter den am häufigsten berichteten Problemen auf.

Insgesamt ist es daher nicht verwunderlich, dass die Vorhersagekraft für den Messparameter *Aggression gegenüber Menschen* nach Studie, Art der Verhaltensbeurteilung und Ziel der Aggression in den untersuchten Publikationen variiert.

In den meisten Wesenstests konnten lediglich die negativen prädiktiven Werte ermittelt werden, da die positiv für *Aggression* getesteten Hunde häufig euthanasiert wurden. Eine

vorsichtige Vermittlung wäre hier mindestens aus Studienzwecken interessant, um die tatsächliche Prävalenz für *Aggression* im neuen Zuhause zu ermitteln.

So haben CHRISTENSEN et al. (2007) im *Assess-A-Pet* Wesenstest einen negativen prädiktiven Wert für *Aggression gegenüber BesitzerInnen* von 95.5 %, für *territoriale Aggression* von 48.5 % und für *Beuteaggression* von 84% ermittelt. Die Wahrscheinlichkeit, dass ein negativ getesteter Hund im neuen Zuhause auch tatsächlich keine mittel- oder hochgradige Aggression zeigt, liegt bei nur 59.1%. Es gibt demnach eine hohe Zahl falsch-negativ getesteter Hunde, etwa vier von zehn Hunden würden trotz bestandenem Wesenstest dennoch mittel- bis hochgradige Aggression im neuen Zuhause zeigen. BOLLEN et HOROWITZ (2008) haben für den *Assess-A-Pet* Wesenstest einen negativen prädiktiven Wert für *Aggression* im Allgemeinen von 92 % ermittelt, MORNEMENT et al. (2015) im *B.A.R.K.* Wesenstest einen negativen prädiktiven Wert von 75% und keine signifikante Übereinstimmung. Bei CLAY et al. (2020a) im Wesenstest der *RSPCA* konnte *Aggression* ebenfalls nicht vorausgesagt werden. Hunde, die durch BOLLEN et HOROWITZ (2008) als „*grenzwertig*“ klassifiziert wurden, waren assoziiert mit Rückgabe ins Tierheim aufgrund von aggressivem Verhalten. Wesenstests konnten die Rückgaberate von 19% auf 14% reduzieren innerhalb der zweijährigen Studie. Rückgaben aufgrund von Aggression konnten von 5% auf 3.5% reduziert werden mit viel weniger Vorfällen von ernsthafter Aggression.

Lediglich in der Studie von VAN DER BORG et al. (1991) wurden aggressive Hunde nicht euthanasiert und der Vorhersagewert gemessen – der negative prädiktive Wert liegt hier bei 85 %. Klare Tendenzen lassen sich hier nicht erkennen, da dieser negative prädiktive Wert ohne Euthanasie der Positivgetesteten beispielsweise niedriger ist als bei *Assess-A-Pet* (BOLLEN et HOROWITZ; 92%), aber höher als bei *B.A.R.K.* (MORNEMENT et al., 75 %). Der positive prädiktive Wert lässt sich nicht vergleichen, da durch die Euthanasie der Positivgetesteten diese in der Follow-Up-Studie nicht beurteilt werden können.

Als Gründe für die schlechten Vorhersagewerte in manchen Kategorien werden Schlafmangel, (ggf. okkulte) Krankheiten, Lärm, sozialer und emotionaler Stress und andere hemmende Faktoren während des Tests genannt (CHRISTENSEN et al., 2007). Auch das Fehlen des richtigen Auslösers im Test (soziale Bindung für Aggression gegenüber dem/der BesitzerIn, zu verteidigender Platz für territoriale Aggression, Stimulation einer Jagdsequenz für Beuteaggression) könnte ursächlich sein. Ein großes Problem ist, dass die meisten

Wesenstests lediglich Aggression gegenüber fremden Personen messen (können), obwohl Aggression gegenüber der Besitzerin/ dem Besitzer mitunter die wichtigsten Beschwerden bei adoptierten Hunden verursachen (KIS et al., 2014). Dem hohen negativen prädiktiven Wert für *Aggression gegenüber BesitzerInnen* (95.5%) im *Assess-A-Pet* Wesenstest (CHRISTENSEN et al., 2007) könnte man die Information entnehmen, dass *Aggression gegenüber BesitzerInnen* entweder insgesamt einer niedrigen Prävalenz unterliegt oder gut vorausgesagt werden kann. Aufgrund der Abwesenheit von BesitzerInnen/ Bezugspersonen im Wesenstest erscheint Letzteres eher unwahrscheinlich, der Wesenstest dürfte zur Evaluierung dieses Aspektes daher ungeeignet sein.

Ähnlich dürfte es bei *Beuteaggression* und *territorialer Aggression* sein – Umstände, die im Wesenstest ebenfalls kaum simulierbar sind. Ein allgemeiner Kritikpunkt von Wesenstests ist die fehlende „natürliche Umgebung“ des Hundes abseits seines Zwingers. Territoriales Verhalten kann daher beispielsweise nur begrenzt getestet werden (VAN DER BORG et al., 1991). Aufgrund der hohen Zahl der Falsch-Negativen in einigen Messparametern ist der negative prädiktive Wert bei *territorialer Aggression* absolut unbefriedigend (CHRISTENSEN et al., 2007). Die Wahrscheinlichkeit, dass ein negativ getesteter Hund im neuen Zuhause keine territoriale Aggression in irgendeiner Art (inkl. Bellen) zeigt, liegt bei nur 48.5%, da kann man gleich eine Münze werfen. Da territoriale Aggression unter Umständen auch das Verhalten gegenüber Fremden und BesucherInnen beeinflusst, sind schlechte Ergebnisse bezüglich des Vorhersagewertes dieser Situationen ebenfalls nicht verwunderlich.

Die Umsetzung mancher Stimuli ist ebenfalls nur limitiert oder gar nicht möglich, beispielsweise ist die Präsentation einer Katze im Käfig als Auslöser sowohl aus Stressgründen der Katze gegenüber als auch durch das Fehlen der möglichen Jagdsequenz als Test fraglich (CHRISTENSEN et al., 2007), ähnliche Argumente gelten auch für die Evaluierung von Verhalten gegenüber Kindern.

Immer wieder wurde überlegt, ob der professionelle Umgang durch das Tierheimpersonal ein Grund für falsch-negative Testergebnisse bezüglich *Aggression* sein könne, im Vergleich zum Umgang der NeubesitzerInnen (MORNEMENT et al., 2015). Diese können komplette AnfängerInnen sein bezüglich ihres Wissens über kanine Körpersprache und Verhalten. Viele Problemverhalten könnten im Umgang durch erfahrene PflegerInnen seltener oder gar nicht auftreten. CHRISTENSEN et al. (2007) erwähnen weiters die sogenannte „*Honeymoon-*



*Periode*“, eine Zeit unbestimmter Länge, in der ein Hund vorheriges aggressives Verhalten (noch) nicht zeigt. Außerdem ist unklar, ob manche NeubesitzerInnen eventuell Aggression gegenüber sich selbst verschwiegen haben, aus Angst vor Konsequenzen für den Hund oder aus Scham. Dies sind Gründe, aus denen die wahre Prävalenz für Aggression noch größer sein könnte.

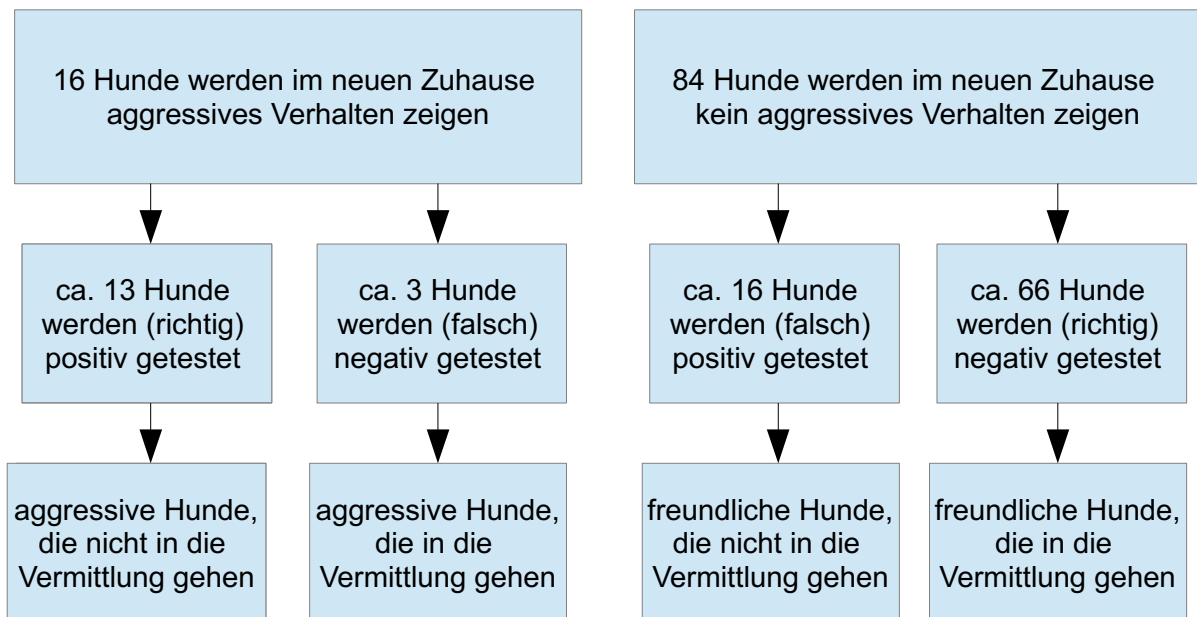
Wie bereits erwähnt, ist keiner der Studien ein realistischer Wert für die Sensitivität und die Spezifität zu entnehmen, da meist jene Individuen euthanasiert wurden, die Aggression im Wesenstest zeigten. Allerdings kann den Daten zum Wesenstest von VAN DER BORG et al. (1991) eine Vierfeldertafel entnommen werden und somit Sensitivität und Spezifität berechnet werden, siehe Tabelle 36.

Tab. 36: Vierfeldertafel zu *Aggression gegenüber Erwachsenen* (VAN DER BORG et., 1991)

<b>Ergebnis der Verhaltensbeurteilung im Tierheim</b>	<b>Verhalten nach der Adoption</b>	
	<b>Hund zeigte Aggression gegenüber Erwachsenen n = 28</b>	<b>Hund zeigt keine Aggression gegenüber Erwachsenen n = 36</b>
<b>Testergebnis positiv n = 31</b>	<b>tatsächlich positive Hunde (a)</b> n = 23	<b>Falsch-positive Hunde (b)</b> n = 8
<b>Testergebnis negativ n = 33</b>	<b>Falsch-negative Hunde (c)</b> n = 5	<b>tatsächlich negative Hunde (d)</b> n = 28

Die Sensitivität ergibt in dieser Studie 82.1%, die Spezifität ergibt 77.8%, der negative prädiktive Wert ergibt 84.8% und der positive prädiktive Wert ergibt 74.2%. In dieser getesteten Population scheint die Prävalenz bei 43.8% für das Zeigen von Aggression gegenüber Erwachsenen zu liegen. Daher ergeben sich die hohen prädiktiven Werte. In dieser Rechnung wären 15.2% der Hunde falsch-negativ. Es stellt sich die Frage, weswegen die Prävalenz in der Studie derart hoch zu sein scheint. Die AutorInnen zählen bereits Knurren und Piloerektion zu aggressivem Verhalten. Außerdem könnte der provokative Testcharakter derartige Ergebnisse initiieren.

Rechnen wir daher mit den Prävalenzdaten von PATRONEK und BRADLEY (2016), das heißt, mit einer gegebenen Prävalenz von 16%, einer berechneten Sensitivität von 82%, einer berechneten Spezifität von 78% und einer Anzahl von 100 im Tierheim getesteten Hunden, ergibt sich folgendes Gedankenexperiment (siehe Abb. 2):



Resultat: Von 10 Hunden, die positiv auf Aggression getestet wurden und dadurch nicht in die Vermittlung gehen, sind 6 Hunde eigentlich nicht aggressiv.

Abb. 2, Gedankenexperiment nach PATRONEK und BRADLEY (2016)

Gehen wir beispielsweise von einer Prävalenz von nur 11 % aus (unter derselben Sensitivität und Spezifität), sind von zehn Hunden, die positiv auf Aggression getestet wurden und dadurch nicht in die Vermittlung gehen, sogar sieben Hunde eigentlich nicht aggressiv. Genau diese Falsch-Positiven sind jedoch die Geschädigten der Wesenstests, da ihnen die Möglichkeit auf Adoption verwehrt wird oder sie gar euthanasiert werden. Bei einer Prävalenz von 16% bzw. 11% lägen die prädiktiven Werte bei 44.8% bzw. 31% (PPW) und 95.7% bzw. 97.2% (NPW). Die Aussagekraft eines positiven Testergebnisses beim Test auf Aggression erscheint also mehr als fraglich. Eine Euthanasieentscheidung aufgrund eines solchen Testergebnisses zu treffen, wäre sinnfrei (PATRONEK et BRADLEY, 2016).

An dieser Stelle stellt sich erneut die Frage, wie die prognostische Validität für *Aggression* in Studien an Hunden in Privatbesitz ausfällt. Diese wurde unter anderem für den Wesenstest

nach KLAUSZ *et al.* untersucht (KLAUSZ *et al.*, 2013). Die Korrelationen zwischen dem Verhalten im Test und den Berichten der BesitzerInnen wurden als stark bezeichnet. Die drei Einzeltests *freundliche Begrüßung*, *Wegnehmen eines Knochens* und *bedrohliche Annäherung* wurden weiters als effektiv darin beschrieben, spezifisches Verhalten entsprechend der Beißvorgeschichte hervorzurufen. Die angegebenen Sensitivitäten in der Studie sind dennoch teilweise sehr gering (12% für *freundliche Begrüßung*, 48% für *Wegnehmen eines Knochens* und 56% für *bedrohliche Annäherung*). In meinen Augen stellt daher diese Studie keine zufriedenstellende Referenzvalidität dar. In einer anderen Studie (BENNETT *et al.*, 2012) wurden die prognostische Validität des SAFER™ Wesenstests und des Assess-A-Pet Wesenstests (allerdings eine modifizierte Version) gemessen. Als Referenzwert galten die Ergebnisse einer Befragung der BesitzerInnen via C-BARQ-Fragebogen. Im Wesenstest waren die BesitzerInnen, analog zum Ablauf in einer Tierheimsituation, nicht anwesend. Die Sensitivität des SAFER™ Wesenstests, *Aggression* festzustellen lag bei 60%, die Spezifität bei 50%. Die Sensitivität des Assess-A-Pet Wesenstests, *Aggression* festzustellen lag bei 73%, die Spezifität bei 59%. Wenn die Ergebnisse des Wesenstests in Unterkategorien aufgeteilt wurden bezüglich des Grades an *Aggression*, war die Validität dennoch schwach und erreichte nicht einmal eine signifikante Übereinstimmung. Die Ergebnisse sprechen nicht für einen hohen Grad an Validität, weswegen auch hier nicht von einer Referenzvalidität ausgegangen werden kann.

Im Wesenstest nach NETTO *et PLANTA* (1997) bei Hunden in Privatbesitz war das Ziel, besonders aggressive Individuen zu identifizieren. Die Hunde wurden in zwei Gruppen eingeteilt (*potentiell aggressive Rassen vs. Kontrollgruppe*), als Referenzwert diente ein ausgefüllter Fragebogen. Hunde mit und ohne Aggressionsvorgeschichte (beider Kontrollgruppen) unterschieden sich signifikant in ihrem Beiß-/ Attackierverhalten während des Wesenstests. Damit wurde der Wesenstest als valide befunden, Hunde mit aggressiven Tendenzen zu detektieren. IVERSEN (2008) hat den Wesenstest *nach NETTO et PLANTA* teilweise modifiziert und angewandt. Als Referenzwert diente auch hier ein Fragebogen, welcher durch die BesitzerInnen ausgefüllt wurde. Die Übereinstimmung von aggressivem Verhalten mit dem Vorbericht lag bei 86.5%. Hunde mit einer Aggressionsvorgeschichte zeigten signifikant stärker aggressives Verhalten im Wesenstest. Dennoch gab es 61% Falsch-Positive, welche im Wesenstest aggressives Verhalten zeigten, dies im Alltag durch den Fragebogen allerdings nicht berichtet wurde. In dieser Studie kam die Autorin zu dem

Ergebnis, dass der niederländische Aggressionstest *nach NETTO et PLANTA* ein hohes Potential aufweist, aggressives Verhalten auszulösen. Siebzig Prozent der Hunde zeigten im Test mindestens einmal aggressives Verhalten und ein Drittel der Hunde biss mindestens einmal zu. Der Test scheint außerordentlich provokativ zu sein und könnte die Prävalenz von Aggression in der getesteten Population daher nicht wahrheitsgemäß wiedergeben. THIESSEN-MOUSSA et al. (2018) haben Wesenstests (angelehnt an eine modifizierte Variante des Wesenstests *nach NETTO et PLANTA*) retrospektiv ausgewertet. Als Referenzwert galt die Einstufung der Hunde als *gefährlich* aufgrund eines Beißvorfalls mit einem Menschen oder mit einem anderen Tier in der Vergangenheit (die Teilnahme an einem Wesenstest ist im deutschen Bundesland Niedersachsen bei als *gefährlich* eingestuften Hunden Vorschrift). Jene Hunde, die als gefährlich eingestuft wurden, weil sie einen Menschen gebissen hatten, zeigten im Wesenstest signifikant häufiger aggressives Verhalten in den Testsituationen, in denen Hund-Mensch-Kontakt bestand und bestanden den Wesenstest häufiger nicht.

Im *Socially Acceptable Behavior test (SAB test)* stellt ebenfalls *Aggression* den wesentlichen Messparameter dar. Während des Wesenstests wird der Hund durch die/den BesitzerIn geführt (PLANTA et DE MEESTER, 2007). Die Übereinstimmung zwischen einer Beißvorgeschichte und aggressivem Beißen während des SAB-Tests lag bei 82%, wenn jegliches aggressives Verhalten mit Beißintention als inakzeptabel galt. Die Sensitivität lag bei 84%, die Spezifität bei 81%, der PPW bei 64% und der NPW bei 93%. Galt das einmalige Zeigen von aggressivem Verhalten mit Beißintention in acht Einzeltests als noch akzeptabel und der Wesenstest somit noch als negativ, stieg die Übereinstimmung auf 87%, die falsch-negativen Hunde zeigten nur territoriale Aggression. Die Sensitivität liegt in diesem Fall bei 67%, die Spezifität bei 95%, der PPW bei 83% und der NPW bei 88%. In einer weiteren Untersuchung wurde die prognostische Validität des Tests untersucht, indem die Ergebnisse des *SAB tests* mit dem Verhalten innerhalb eines Jahres nach dem Wesenstest (berichtet durch die BesitzerInnen) verglichen wurden. Die Vorhersagekraft für aggressives Verhalten mit Beißintention wurde mit 81.8% angegeben. Die Sensitivität liegt bei 69.4% und die Spezifität bei 84.2%. Der Test soll demnach das Auftreten von aggressivem Verhalten mit Beißintention gegenüber Fremden in einem nicht-territorialen Kontext signifikant voraussagen können. Im Vergleich zur Tierheimsituation konnte somit bei Hunden in Privatbesitz eine bessere Übereinstimmung zwischen Wesenstest und Reevaluierung nach einem Jahr festgestellt

werden. Eine weitere Studie (VAN DER BORG et al., 2010) untersucht die Fähigkeit des *SAB tests*, Aggression gegenüber Menschen zu identifizieren. Auch hier dienten Berichte der BesitzerInnen als Referenzwert. Ein Hund galt als aggressiv, wenn er laut BesitzerInnen mindestens einmal einen Menschen gebissen hat bzw. mindestens einmal im Wesenstest einen Menschen attackiert hat. Die Sensitivität des *SAB tests* für die Detektion von Aggression gegenüber Menschen lag bei 0.33 und die Spezifität bei 0.81. Wurde der Grenzwert für einen positiven Test auf das Zeigen von mindestens zwei Attacken erhöht, konnte damit die Spezifität auf 0.93 gesteigert werden. DE MEESTER (2011) haben die Körperhaltungen und Verhaltensstrategien von Hunden im *SAB Wesenstest* analysiert. Die signifikante Korrelation zwischen den *SAB*-Scores in den unterschiedlichen, aufeinander folgenden Einzeltests zeigte, dass die Körperhaltung neben der Art des Stimulus auch durch den vorhergehenden Einzeltest beeinflusst wurde. Die meisten Hunde, die die Puppe oder die *Fake-Hand* bissen, hatten in den spezifischen Einzeltests eine niedrige Haltung, was darauf hinweist, dass die Emotion hinter diesem Verhalten *Angst* war. Aufgrund der niedrigen Vorhersagewerte konnte festgestellt werden, dass es nicht möglich war, alle Informationen bezüglich des Biss- und/oder Fluchtverhaltens später im Test zu erhalten, sondern nur aus den Verhaltensstrategien oder *SAB*-Scores der Hunde in den Einzeltests. DALLA VILLA et al. (2017) haben den *SAB test* in Italien angewandt und wiederum anhand einer Befragung via *C-BARQ* Fragebogen validiert. Hunde, welche im Wesenstest aggressive Reaktionen gezeigt haben, hatten im Fragebogen signifikant höhere („schlechtere“) Punktezahlen hinsichtlich *Aggression gegenüber Fremden/ gegenüber der/dem BesitzerIn* und *gegenüber einem vertrauten Hund*. Der *SAB test* wurde damit als valide für die Voraussage von *Aggression gegenüber Fremden* erklärt. Im Wesenstest *DMA (Dog Mentality Assessment)* bei Hunden in Privatbesitz waren die Übereinstimmungen zwischen dem Wesenstest und einer Follow-Up-Befragung via *C-BARQ* ein bis zwei Jahre später im Messparameter *Aggressivität* eher schwach (0.11 bis 0.15) (SVARTBERG, 2005).

Aus diesen Ergebnissen geht hervor, dass Wesenstests bei Hunden in Privatbesitz zwar signifikante und teils hohe Übereinstimmungen zu Vorberichten etc. erreichen, aber auch, dass dabei die Zahl der falsch-positiven Hunde nicht zu vernachlässigen ist – so sind 41% der Hunde in *Assess-A-Pet* (BENNETT et al., 2012) falsch-positiv und 50% in *SAFER™*. Im *SAB*-Test fällt die Zahl zwar geringer aus (5-19%), aber die Sensitivität ist mit 33% (VAN

DER BORG et al., 2010), bzw. 67% (PLANTA et DE MEESTER, 2007) immer noch zu gering, um meiner Meinung nach eine angemessene Referenzvalidität zu gewährleisten.

#### 4.2.3.1.2 **Angst und Furcht**

Auch hier ist festzuhalten, dass in einigen Studien/ Tierheimen „*extrem ängstliche*“ Hunde euthanasiert wurden (beispielsweise CLAY et al., 2020a; POULSEN et al., 2010) und dadurch wie auch bei Aggression die untersuchte Population verändert wurde.

Bei Wesenstest differierten die Aussagen zur Vorhersagekraft von Furcht und Angst sehr stark. So konnte laut CLAY et al. (2020a) *Angst gegenüber den BesitzerInnen und/oder Kindern* basierend auf dem Wesenstest der RSPCA signifikant vorausgesagt werden. Auch MORNEMENT et al. ermittelten 2014 für den B.A.R.K. Wesenstest eine moderate statistisch signifikante Übereinstimmung und auch 2015 eine signifikante Übereinstimmung im Bereich *Angst gegenüber Fremden*. Die moderate statistisch signifikante Übereinstimmung im B.A.R.K. Wesenstest bei MORNEMENT et al. (2014) traf jedoch nur auf *Furcht* zu, nicht auf *Angst*.

Beispielhaft ist hier besonders der sogenannte „*Regenschirm-Test*“ zu nennen, dieser löst starke furchtsame Reaktionen aus (TAYLOR et MILLS, 2006). Man kann diskutieren, ob es sich dabei um eine normale Alltagssituation handelt und somit dessen Aussagekraft überhaupt relevant ist.

Gerade bei *Angst* halte ich eine Vorhersage über das tatsächliche Auftreten und die Entwicklung für sehr vage. Ob Ängste bestehen und wie sich ein ängstlicher Hund entwickelt, hängt stark vom Umgang und den Gegebenheiten ab. Manche Hunde sind von ihrer Grundpersönlichkeit ein ängstlicher Typ und manche Hunde reagieren auf die neue, für sie unberechenbare Tierheimsituation mit temporärer Ängstlichkeit und/oder daraus resultierender Selbstverteidigung (siehe *Aggression*) – die Differenzierung im Tierheim hierzu bedarf weiterer Forschung. Hier könnten Fragebögen der VorbesitzerInnen zur individuellen Temperamenteinschätzung von Hunden, wie zum Beispiel dem PANAS (SHEPPARD et MILLS, 2002) oder dem DIAS (WRIGHT et al., 2011) hilfreich sein. Hunde, welche nur auf die Tierheimsituation ängstlich reagieren, haben großes Potenzial, in einem neuen Zuhause zeitnah zu ihrem ursprünglichen Selbstbewusstsein zurückzukehren, während dies bei Hunden mit einem ängstlichen Temperament nicht so schnell der Fall sein wird.

Vielversprechend ist in jedem Fall eine umfangreiche Einschulung über den Umgang mit ängstlichen Hunden für InteressentInnen all jener Hunde, welche im Tierheim ängstliches Verhalten zeigen. Eine Zusammenarbeit mit kompetenten TrainerInnen post-Adoption sollte die Voraussetzung für eine Vermittlung darstellen. Auch die Gegebenheiten sollten zu einem ängstlichen Hund passen, so ist beispielsweise von einem neuen Zuhause in der Innenstadt abzusehen.

MEYER et FORKMAN (2014) haben die Ergebnisse des *DMA Wesenstest (Dog Mentality Assessment)* mit denen eines später beantworteten Fragebogens bei Hunden in Privatbesitz verglichen, um den Effekt von Eigenschaften von Hund und Mensch auf die Beziehung zueinander zu untersuchen. Als einzige Eigenschaften, die die Mensch-Hund-Beziehung vorhersagen können, ergaben sich *Angst* und *angstbezogene Verhaltensprobleme*. SVARTBERG (2021) hat den Wesenstest *Assessment of Behavior in Canines/ Behaviour and Personality Assessment in Dogs* aufgegriffen und die Validität untersucht. Als Referenzwert wurden Befragungen durch einen Fragebogen verwendet. Signifikante Korrelationen konnten besonders bei den Messparametern *Angst vor Fremden* und *Interesse an Fremden* im Fragebogen erreicht werden (beispielsweise war Umgänglichkeit im Wesenstest mit -0.40 negativ korreliert mit *Angst vor Fremden* im Fragebogen und mit 0.43 positiv korreliert mit *Interesse an Fremden* im Fragebogen). Diese Ergebnisse sprechen dafür, dass Angst auch bei Hunden in Privatbesitz bessere Validität erreicht als beispielsweise *Aggression*.

#### **4.2.3.1.3 Trennungsangst**

Trennungsangst ist ein gängiger Abgabegrund für Hunde im Tierheim und auch ein häufiges Problem nach der Adoption. Leidet ein Hund unter starker Trennungsangst, mindert dies seine Vermittlungschancen sehr stark. Selbst pensionierte NeubesitzerInnen sind darauf angewiesen, den Hund das ein oder andere Mal für einen Arztbesuch oder einen Einkauf allein zu lassen. Hunde, die nicht allein bleiben können, erfordern im Alltag einen hohen Aufwand an Training oder Management. Auch sind die Möglichkeiten, dies bereits vor der Adoption im Tierheim zu trainieren eingeschränkt.

Hunde, die unter Trennungsangst leiden, zeigen bei Separation von ihren BesitzerInnen oder Menschen im Allgemeinen Zeichen von Verzweiflung, darunter Vokalisation, Zerstörungsverhalten, Unsauberkeit, Speicheln, Ausbrechen und Depression (CLAY et al., 2020c).

Um Trennungsangst zu testen, wurde in den Wesenstests der Hund beispielsweise in einem ihm fremden Raum für einige Zeit allein gelassen. Die Sensitivität lag in den Wesenstests für das Erkennen von Trennungsangst bei VAN DER BORG et al. (1991) bei 83% und der negative prädiktive Wert bei 92%. Bei CLAY et al. (2020a) konnte Trennungsangst im Wesenstest der *RSPCA* dagegen nicht vorausgesagt werden.

Wurde ein Testinstrument bereits in einer vergleichbaren Population zu einem hohen Grad validiert, kann dies ebenfalls eine Aussage als Referenzvalidität in der untersuchten Population treffen (PATRONEK et al., 2019). DÖRING et al. (2017b) haben die Vorhersagekraft eines Wesenstests bei Laborbeagles gemessen, indem sie Hunde in der Forschungseinrichtung getestet haben und eine telefonische Follow-Up-Befragung jeweils eine und zwölf Wochen nach der Vergabe in Privatbesitz durchführten. Die Interviews ergaben eine Veränderung zugunsten erwünschter Verhaltensweisen innerhalb der elf Wochen und daher nur eine niedrige bis moderate Übereinstimmung. Als Hauptprobleme wurden Trennungsangst, Zerstören von Objekten und mangelnde Stubenreinheit genannt. Die Indikation dafür, Laborhunde zu testen, ähnelt am ehesten einer Testung von Tierheimhunden, da auch hier das Verhalten nach Vergabe an private Plätze im Vordergrund steht und eine Vermittlung ebenfalls gegen die Alternative einer Euthanasie steht. Dennoch halte ich die beiden Populationen für nicht ausreichend vergleichbar, da die Vergangenheit und die Erfahrungen der individuellen Hunde stark differieren. Die aktuelle Lebenssituation weist zwar Parallelen auf, aber ein Laborhund kennt in der Regel ausschließlich dieses (oft eingeschränkte) Leben, während ein Tierheimhund verschiedenste Vorgeschichten haben kann. Auch in den beschriebenen Verhaltensproblemen nach der Adoption sehe ich Differenzen zu der Tierheimhundepopulation, da diese Probleme hauptsächlich daraus resultieren könnten, dass die Laborhunde mit dem Leben in einer häuslichen Umgebung zumeist (noch) nicht vertraut sind. Tierheimhunde dagegen kennen meist das Leben in einer Familie bereits zu einem gewissen Grad. In der Beurteilung von Tierheimhunden liegt das Hauptaugenmerk dagegen auf der Detektion diverser Formen von Aggression, während dies bei Laborhunden, welche gezielt auf die Ausprägung einer hohen Toleranzschwelle (beispielsweise bei Manipulation) gezüchtet wurden, eine untergeordnete Rolle spielen könnte.



#### 4.2.3.1.4 **Freundlichkeit**

Im Gegensatz zu *Aggression* und *Angst* war bei allen AutorInnen, welche die Vorhersagekraft von *Freundlichkeit* untersucht haben, eine durchaus statistisch signifikante Übereinstimmung zu finden. In CLAY et al. (2020a) konnte *Freundlichkeit* im Wesenstest der RSPCA signifikant vorausgesagt werden, auch MORNEMENT et al. (2014) stellten im B.A.R.K. Wesenstest eine moderate Übereinstimmung fest und in der Folgestudie (2015) eine ebenfalls signifikante Übereinstimmung bezüglich *Freundlichkeit gegenüber BesucherInnen*.

#### 4.2.3.1.5 **Verträglichkeit mit Artgenossen**

Die Untersuchung der Vorhersagekraft von Verträglichkeit mit Artgenossen fiel insgesamt sehr spärlich aus in den Studien. Auch sind die evaluierten Parameter sehr unterschiedlich. In den Wesenstests ist der Messparameter meist *Aggression gegenüber Artgenossen* (VAN DER BORG et al., 1991; CHRISTENSEN et al., 2007; WEISS, 2002; BENNETT et al., 2015; POULSEN et al., 2010). Im *Temperament Test* wird auf *Umgänglichkeit mit Artgenossen* getestet und SHABELANSKY et al. (2015) haben sowohl *Freundlichkeit*, als auch *Angst* und *Aggression* beurteilt.

Der negative prädiktive Wert für *Aggression gegenüber Artgenossen* lag zwischen 83% (CHRISTENSEN et al., 2007; *Assess-A-Pet*) und 95% (VAN DER BORG et al.). Ein möglicher Grund für die Differenz könnte der Zeitpunkt der Follow-Up-Studie sein (bis zu 13 Monate vs. ein bis zwei Monate). In einer größeren Befragungszeitspanne steigt die Wahrscheinlichkeit für eine Auseinandersetzung mit einem Artgenossen natürlich, da beispielsweise in einem längeren Zeitraum eine größere Diversität an Hunden getroffen werden kann. Die Sensitivität des Wesenstests für das Erkennen von *Aggression gegen Artgenossen* lag bei VAN DER BORG et al. bei 83%.

Aufgrund der starken demographischen Varianz des Stimulus „Artgenosse“ ist innerartliche Verträglichkeit nur limitiert testbar in einem Wesenstest, da keine große Palette diverser Hundetypen präsentiert werden kann. Weiters findet der Test meist aus Sicherheitsgründen ausschließlich an der Leine statt. Es gibt beispielsweise Hunde, die ausschließlich Leinenaggression zeigen. Es gibt auch Hunde, die keine brachycephalen Hunde mögen oder keine großen schwarzen Hunde. Wie aussagekräftig können diese Informationen erhoben werden?

THIESSEN-MOUSSA et al. (2018) haben Wesenstests (angelehnt an eine modifizierte Variante des Wesenstests *nach NETTO et PLANTA*) retrospektiv ausgewertet. Test-situationen mit Hund-Hund-Kontakt schienen dabei nicht ausreichend aussagekräftig zu sein, um einen Beißvorfall mit einem anderen Tier vorauszusagen. Als Referenzwert galt die Einstufung der Hunde als *gefährlich* aufgrund eines Beißvorfalls mit einem Menschen oder mit einem anderen Tier in der Vergangenheit. Somit scheint auch bei Hunden in Privatbesitz keine angemessene Aussagekraft bei der Testung der Verträglichkeit möglich zu sein.

Auch wenn der Stimulus existenziell ähnlich für Menschen erscheint, können sie durch Hunde sehr unterschiedlich wahrgenommen werden durch andere Faktoren im Kontext (PATRONEK et BRADLEY, 2016). Besonders im Teilaspekt *Verträglichkeit* von Verhaltens-evaluierungen spiegelt sich dieses Problem wider. Hier ist wahrscheinlich der Befragung mittels Fragebögen und Beobachtungen durch Ethogramme der Vorzug zu geben, auch in Hinblick auf die Sicherheit und das Wohlergehen aller Beteiligten während eines provozierten Tests in der „realen“ Welt.

Nach Möglichkeit könnte die Haltung von Hunden in Gruppen eine artgerechte Alternative zur Einzelhaltung sein und dem Hund die Möglichkeit bieten, seine sozialen Fertigkeiten anzuwenden und auszubauen. Eine Hypothese könnte lauten, dass Hunde aus Tierheimen mit Gruppenhaltung im neuen Zuhause über eine höhere soziale Kompetenz gegenüber Artgenossen verfügen als Hunde aus Einzelhaltung. Bisher konnte ich keine Studie finden, die diese Hypothese aufgreift und bestätigt/ widerlegt. Die Tatsache, dass ein Hund im Tierheim in einer Gruppe mit Artgenossen gelebt hat, bedeutet aber nicht zwangsläufig, dass dieser in Folge nicht dennoch Leinenaggression zeigen könnte, was durch viele NeubesitzerInnen als Problem gesehen wird.

#### **4.2.3.1.6 Umgänglichkeit mit Kindern**

Kinder sind die häufigsten Opfer von Hundebissen (BOLLEN et HOROWITZ, 2008), allerdings ist es unmöglich, das Verhalten von Kindern in einem Wesenstest authentisch zu simulieren. In einigen Wesenstests wird die Reaktion auf Kinder getestet (z.B. bei *Assess-A-Pet*) und in anderen Wesenstests nicht (z.B. *SAFER™*). Es gestaltet sich sehr schwierig, die Vorhersagekraft dieses Messparameters zu messen. Im Wesenstest ist aus ethischen und aus Sicherheitsgründen ausschließlich die Anwendung einer lebensgroßen Puppe indiziert.

Beim geringsten Zweifel der Freundlichkeit im Wesenstest sollte von einer Vermittlung zu (Klein-)Kindern abgesehen werden. Ein weiteres Problem ist die fehlende Konsistenz im Verhalten von (Klein-)Kindern. Viele kleine Kinder reagieren für den Hund unberechenbar, indem sie beispielsweise plötzlich schreien, hinfallen oder herumfuchteln. Bei jagd-begeisterten Hunden können weglaufende Kinder zu Beuteverhalten führen. Es ist fraglich, ob eine Vermittlung von Hunden mit unbekannter Vergangenheit zu Kleinkindern überhaupt Sinn macht, zumal die Eltern in dieser Zeit ihr Hauptaugenmerk auf die Kleinkinder legen müssen und weniger Ressourcen für die Eingewöhnung eines Hundes haben. STERNBERG (2017) beschreibt diese InteressentInnen als sogenannte *Level 1 adopters*. Zu ihnen zählen alle Familien mit Kindern unter sieben Jahren und sogar junge Menschen mit dem Potential zum Gründen einer Familie. Diese InteressentInnen sollten nur Hunde mit den höchsten Reizschwellen für Aggression jeglicher Art adoptieren.

Die Verwendung einer lebensgroßen Puppe zur Detektion von Aggression gegenüber Kindern hat sich beispielsweise in der Studie von VAN DER BORG et al. (1991) bewährt, wobei auch bei Abwesenheit von Aggression im Test Vorsicht geboten sein sollte, besonders wenn an anderer Stelle im Wesenstest Aggression gezeigt wurde. Genaue Angaben zur Vorhersagekraft lassen sich aber nicht entnehmen.

An dieser Stelle könnten Fragebögen eine wertvolle Stütze darstellen, allerdings wurde in keinem Fragebogen die *Umgänglichkeit mit Kindern* explizit erfragt. Leider ist aufgrund der Inkonsistenz im Verhalten der Kinder vermutlich auch hier kein aussagekräftiger Rückschluss möglich. Die Vorinformation „*hat mit Kindern zusammengelebt*“ ist nicht automatisch eine Referenz, da Probleme in dieser Beziehung u.U. zur Abgabe des Hundes geführt haben könnten.

#### **4.2.3.1.7 Ressourcenverteidigung – Sonderfall Futter**

Besitzergreifendes Verhalten oder ressourcenbedingte Aggression wird auch als Ressourcenverteidigung bezeichnet (MOHAN-GIBBONS et al., 2012). Dies repräsentiert aggressives Verhalten gegenüber Menschen hochwertige Dinge betreffend (dies können Futter, Spielzeug, Schlafplätze, aber auch Menschen selbst sein). Ethologisch lässt sich das Verhalten aus dem Wettbewerb um Ressourcen erklären und wird individuell durch Genetik, Vergangenheit und Verfügbarkeit von Ressourcen beeinflusst. Ressourcenverteidigung kann

auch durch andere Hunde, Kinder und/oder Unterschreitung der Individualdistanz provoziert werden. Die Thematik Futterverteidigung hat dabei besondere Brisanz. Hunde, die Futterverteidigung im Tierheim zeigen, werden weniger adoptiert, werden zum Teil nur mit strengen Auflagen vergeben, haben daher längere Tierheimaufenthalte und werden häufiger euthanasiert als andere Hunde (MOHAN-GIBBONS et al., 2018; MCGUIRE et al., 2019).

Die Prävalenz von Futterverteidigung ist schwierig zu erfassen. Siebzehn Prozent der Hunde, welche durch MOHAN-GIBBONS et al. (2018) mittels eines Wesenstests evaluiert wurden, zeigten Futterverteidigen. Davon wurden allerdings 80% nur durch den Wesenstest als Futterverteidiger identifiziert und nicht durch Beobachtungen des Tierheimpersonals oder Berichte der VorbesitzerInnen. Da nur etwa die Hälfte der Hunde einen Wesenstest durchlaufen hat, könnten Werte von 17% das Vorhandensein von Futteraggression in der Tierheimhundepopulation überrepräsentieren. Offensichtlich leicht-vermittelbare Hunde wurden eventuell gar nicht erst evaluiert. Nachdem der Einzeltest „Futterverteidigung“ im Zuge der Studie ausgeschlossen wurde, wurden nur 3% der Hunde als Futterverteidiger klassifiziert. In der Studie von MOHAN-GIBBONS (2012) gaben 71 von 77 US-amerikanischen Tierheimen an, futterverteidigendes Verhalten zu testen, bei einer durchschnittlich angegebenen Prävalenz von 14% (mit einer Spannweite von 7 bis 30%). Ressourcenverteidigung wurde außerdem mit 71% als Hauptgrund dafür genannt, dass Hunde als unvermittelbar identifiziert werden. Nur 34% der befragten Tierheime gaben an, im Falle von futterverteidigendem Verhalten an ebendiesem zu arbeiten, 51% hingegen unternehmen keinerlei Versuche der Vermittlung in dem Fall. Die Prävalenz für das Auftreten von *Aggression beim Wegnehmen von Spielzeug, Knochen oder anderen hochgeschätzten Objekten* lag bei MCGUIRE et al. (2020) zwischen 14.4% (*Assess-A-Pet* Wesenstest sämtlicher Hunde, unabhängig vom Vorbericht) und 29.5% (VorbesitzerInnen).

In den Wesenstests lag die Sensitivität für das Erkennen von Futterverteidigung zwischen 39% (MARDER et al., 2013; *Match-Up*) und 67% (MCGUIRE et al., 2020; *Assess-A-Pet*). Der positive prädiktive Wert lässt sich bei MOHAN-GIBBONS et al. (2012) mit 10% berechnen und bewegt sich maximal bis hin zu 55% bei MARDER et al. (2013; *Match-Up*). Der negative prädiktive Wert für das Verteidigen von Futter lag zwischen 78% (MARDER et al., 2013; *Match Up*) und 97.5% (MCGUIRE et al., 2020; *Assess-A-Pet*) in Studien ohne Euthanasie von Positiven. In Studien, in denen Positiv-Getestete euthanasiert wurden, lag

der negative prädiktive Wert zwischen 91% (CLAY et al., 2020a; Wesenstest der *RSPCA*) und 95.5% (CHRISTENSEN et al., 2007; *Assess-A-Pet*). Der Wert ist somit nicht offensichtlich viel besser. Der positive prädiktive Wert war für alle publizierten Analysen niedrig, denn über die Hälfte der als Ressourcenverteidiger identifizierten Hunde zeigte kein derartiges Verhalten im neuen Zuhause. Es stellt sich also wiederholt die Frage, inwiefern Ressourcenverteidigung sich nach der Adoption tatsächlich als problematisch erweist, bzw. deswegen auch zur Rückgabe des Hundes ins Tierheim führt.

Besonders brisant ist daher in diesem Fall, dass Tierheime, die Nachkontrollen von vergebenen futterverteidigenden Hunden gemacht haben, angaben, dass bei diesen kaum Probleme aufgrund von Futterverteidigung durch die neuen BesitzerInnen berichtet wurden. Somit ist die Diagnose „futterverteidigender Hund“ durch Verhaltensbeurteilungen lebensentscheidend, da noch immer in vielen Tierheimen diese Hunde euthanasiert werden. Aggression beim Wegnehmen von Futter oder eines Knochens wurde zum Beispiel in der Studie von VAN DER BORG et al. (1991) von niemandem als Problemverhalten angegeben, ein Viertel der NeubesitzerInnen interagierten jedoch einfach nicht mit dem Hund, während er gefressen hat. Aggression gegenüber Erwachsenen und Trennungsangst führten dagegen am häufigsten zur Rückgabe des Hundes. Nur einer von 25 Hunden mit hochgradiger Ressourcenverteidigung (evaluiert via *Assess-A-Pet*) hat im neuen Zuhause gebissen und wurde euthanasiert, alle anderen wurden (teilweise mehrmals) ins Tierheim zurückgegeben, aber letztendlich erfolgreich vermittelt (MCGUIRE, 2019). Bei keinem dieser Hunde wurde im Abgabebogen „Ressourcenverteidigung“ oder andere Formen von Aggression als Grund der Rückgabe aufgeführt. Auch in der Studie von MCGUIRE et al. (2020; *Assess-A-Pet*) empfand keiner der NeubesitzerInnen von Positiv-Getesteten das Verteidigen als gravierendes Problem. Interessant ist auch, dass die meisten NeubesitzerInnen angaben (auch die der Hunde mit Futteraggression), dass sie den gleichen Hund noch einmal adoptieren würden, unabhängig davon ob er Futteraggression zeigt oder nicht (MARDER et al., 2013; *Match Up*). Die NeubesitzerInnen der Hunde in der Studie von MOHAN-GIBBONS (2012, *SAFER™*) berichteten über eine starke Beziehung zu ihren Hunden, auch hier wurde kein Hund aufgrund von Futterverteidigen zurückgegeben. Das Verteidigen von Futter wurde bei positiv Getesteten drei Wochen nach der Vermittlung selten berichtet, nach drei Monaten gar nicht.

Generell ist zu sagen, dass die Abwesenheit von Futteraggression im Wesenstest informativer ist als dessen Anwesenheit. Daher spricht der hohe NPW für eine Sinnhaftigkeit im Rahmen einer Befragung der VorbesitzerInnen – aber nur, um zu wissen, dass dieses Problemverhalten mit hoher Wahrscheinlichkeit nicht auftritt.

Die Untersuchung von MOHAN-GIBBONS et al. (2018) ergab, dass nach dem Ausschluss des Einzeltests *Futterverteidigung* die Zahl der Bisse und Verletzungen von Tierheimpersonal, Freiwilligen und NeubesitzerInnen nicht angestiegen ist. Die Zahl der Adoptionen ist sogar gestiegen, da weniger Hunde als Futterverteidiger identifiziert wurden. Es gab dennoch keinen signifikanten Unterschied bezüglich der Rückgaberate, ebenso wenig wie bezüglich der Melderate von Vorfällen nach der Adoption. Nur 17% der Hunde mit Futterverteidigung zeigten hochgradiges Futterverteidigen – dieser Wert war vor und nach Ausschluss des Einzeltests *Futterverteidigung* gleich hoch. Dies zeigt, dass für die Identifikation von schwerer Futteraggression der Wesenstest nicht zwingend notwendig zu sein scheint.

In einem Punkt sind sich alle Studien einig: das Testen von Ressourcenverteidigung im Wesenstest ergibt keinen Vorteil. Zusammenfassend kann nicht mehr empfohlen werden, diesen Einzeltest im Rahmen von Wesenstests durchzuführen. Er scheint die Sicherheit für das Tierheimpersonal und die NeubesitzerInnen nicht zu beeinflussen. Stattdessen sollten nach Möglichkeit Informationen diesbezüglich von VorbesitzerInnen eingeholt werden und der Hund durch das Tierheimpersonal in der täglichen Routine beobachtet werden, beispielsweise beim Wegräumen der leeren Schüssel. Im Wesenstest der *Washington State University* wird anstatt der *Fake-Hand* alternativ besseres Futter (beispielsweise Dosenfutter auf einem Holzspatel) angeboten, was weit weniger provokativ ist. Weiters kann jeder Hund Futteraggression zeigen, selbst wenn er im Tierheim nicht durch dieses Problem aufgefallen ist. NeubesitzerInnen sollten jedenfalls gut aufgeklärt werden über bisherige Auffälligkeiten und allgemeine Bedürfnisse und Verhaltensweisen von Hunden.

#### **4.2.3.1.8 Vermittelbarkeit**

Im Entscheidungsprozess von AdoptantInnen spielen viele Faktoren eine Rolle, beispielsweise die Größe des Hundes, der Felltyp, das Auftreten, die Persönlichkeit und das Verhalten gegenüber der/dem InteressentIn (CLAY et al., 2020c). COHEN et TODD (2019) haben Präferenzen der Hundewahl von TierheimbesucherInnen und AdoptantInnen unter-

sucht. Nach Freundlichkeit und Verspieltheit standen das Alter und die Größe auf der Liste der Prioritäten. So bevorzugten beispielsweise 65% der BesucherInnen einen Hund unter zwei Jahren und 59% einen mittelgroßen Hund. In einigen Studien stand daher nicht die Verhaltensbeurteilung an sich im Vordergrund, sondern die Beurteilung der individuellen Vermittelbarkeit des Hundes. Zu diesen Studien gehörten FRANK (2007), ZILOCCHI et al. (2015), LUCIDI et al. (2005; *Ethotest*) und STICCO et al. (2011; *Ethotest*).

Die statistischen Validitätskriterien können für „Vermittelbarkeit“ in FRANK (2007) aus den publizierten Daten errechnet werden. Als Referenzwert ist allerdings nur der Ist-Zustand nach vier Wochen möglich (*Hund vermittelt/ nicht vermittelt*), spätere Vermittlungen oder Rückgaben können nicht beachtet werden. Es errechnen sich hierbei Werte von einer Sensitivität mit 13%, Spezifität mit 85%, NPW mit 67% und ein PPW mit 30%. Der einzige Wert, der akzeptabel ist, wäre also die Spezifität. Ob die Hunde innerhalb der vier Wochen (FRANK, 2007) bereits Trainingsprogramme erhalten haben, ist ebenfalls jedoch unklar – dies könnte die Falsch-Positiven erklären. Die Falsch-Negativen lassen sich durch optische und andere demographische Faktoren erklären, welche neben dem Verhalten ebenfalls eine Rolle bei der Vermittelbarkeit spielen.

Auch aus den Daten von ZILOCCHI et al. (2015) können prognostische Validitätskriterien errechnet werden. Siebzehn der 26 Hunde im Wesenstest nach ZILOCCHI et al. wurden vermittelt, davon 70% (12 Hunde) aus Kategorie 1 (*sofort vermittelbar*), 18% (3 Hunde) aus Kategorie 2 (*kritische Vermittelbarkeit*) und 12% (2 Hunde) aus Kategorie 3 (*schwer vermittelbar*). Interessant finde ich - sehen wir als Messparameter nur „Unvermittelbarkeit ja/nein“ an, dass 20% der Hunde in Kategorie 1 dennoch nicht vermittelt wurden, der negative prädiktive Wert für Unvermittelbarkeit wäre hier 80%. Fünfzig Prozent der Hunde wurden vermittelt, trotz dass sie als nicht/schwer vermittelbar (Falsch-Positiv für Unvermittelbarkeit) laut Kategorie 3 beurteilt wurden, der positive prädiktive Wert wäre also hier nur 50%. Betrachten wir als Referenzwert „vermittelt“ (14 Hunde) und „nicht vermittelt“ (5 Hunde), wäre die Sensitivität für die Unvermittelbarkeit (nicht vermittelt und als unvermittelbar erkannt) bei 40% (2 aus Kategorie 3 nicht vermittelte Hunde/ 5 nicht vermittelte Hunde) und die Spezifität (vermittelt und als vermittelbar erkannt) bei 86% (12 aus Kategorie 1 vermittelte Hunde / 15 vermittelte Hunde).

In beiden Studien zeigt sich, dass ein Testergebnis „vermittelbar“ für gute Vermittlungschancen spricht, ein „unvermittelbar“ dagegen nicht allzu viel Aussagekraft zugesprochen werden kann. Beide Beurteilungsverfahren wurden in Ländern mit einer „no-kill-Politik“ durchgeführt, in anderen Ländern würde die hohe Zahl an Falsch-positiven Hunden (50% in ZILOCCHI et al. und 70% in FRANK) für diese Hunde fatale Folgen haben.

Im Ethotest (LUCIDI et al., 2005) soll zum Einen die Vermittelbarkeit gesteigert und zum anderen, besonders durch die adaptierte Version von STICCO et al. (2011) auch beurteilt werden. Die Sensitivität für Vermittelbarkeit wird hier mit 100% von den AutorInnen angegeben, wie jedoch bereits in den Ergebnissen angedeutet, wurde hier das Nichtvorhandensein von vorgegebenen unerwünschten Verhaltensweisen bei den NeubesitzerInnen abgefragt. Für eine Beurteilung und Diskussion der statistischen Validitätskriterien sind jedoch leider nur unzureichende Informationen zu den getesteten Hunden vorhanden. Der positive prädiktive Wert für *Unvermittelbarkeit* bewegt sich zwischen 30 und 50%, was aufzeigt, dass ein Hund, welcher zwar als *unvermittelbar* kategorisiert wurde, dennoch Chancen auf Vermittlung hat. Daher halte ich auch diese Versuche, einen Hund zu beurteilen, für nicht zielführend.

#### **4.2.4 Einfluss der Abwesenheit einer Mensch-Hund-Beziehung**

VAN DER BORG et al. (1991) erwähnen als Kritikpunkt der Wesenstests besonders die Abwesenheit einer Hund-HalterInnen-Beziehung. Viele Problemverhalten resultieren aus der Art der Beziehung zwischen Hund und Mensch, und sind daher schwer in Abwesenheit der Bezugsperson zu detektieren. Dennoch konnten beispielsweise für *Trennungsangst* eine hohe Sensitivität und ein hoher negativer prädiktiver Wert erzielt werden. Auslöser ist hier jedoch oft das Alleinbleiben ohne menschlichen Kontakt im Allgemeinen. Nur wenige Parameter, beispielsweise das *Ziehen an der Leine* und *Ungehorsam* scheinen weitestgehend unabhängig vom Besitzer zu sein.

KIS et al. (2014) haben die Ergebnisse der Wesenstests *nach* KLAUSZ et al. von 50 (zufällig ausgewählten) Hunden in Privatbesitz verglichen, in An- und Abwesenheit der Besitzerin/ des Besitzers. Die beiden Tests fanden innerhalb von ein bis drei Wochen statt. Die Studie von KIS et al. (2014) ist zwar nicht direkter Inhalt dieser Arbeit, da es sich nicht um Hunde im Tierheim handelt – das Ziel soll aber dennoch eine Einschätzung sein, ob die Anwesenheit



einer menschlicher Bezugsperson, wie beispielsweise der Pflegerin/des Pflegers, einen Einfluss auf das Ergebnis des Wesenstests nehmen könnte. Die Hunde zeigten in Anwesenheit der Besitzerin/ des Besitzers häufiger Aggression, sowohl im Einzeltest „*Futter wegnehmen*“, als auch im Einzeltest „*Bedrohung*“. *Furcht/Unwohlsein* im Einzeltest „*Annäherung*“ wurde in Anwesenheit der BesitzerInnen seltener gezeigt. Als möglicher Grund wird die sichere Basis angegeben, die eine Bezugsperson dem Hund gibt. Auch in einer anderen Studie (DE MEESTER et al., 2008) wurde demonstriert, dass die An- bzw. Abwesenheit der BesitzerInnen die Ergebnisse eines Wesenstests beeinflussen.

In allen untersuchten Wesenstests sind die prüfenden Personen dem Hund nach Vorgaben des Testablaufs möglichst unbekannt, um ein objektives Ergebnis zu erzielen. Ein/e dem Hund bekannte/r PflegerIn wäre der Situation im neuen Zuhause aber vermutlich ähnlicher. Dennoch könnte das Ergebnis auch hier stark durch die Art der Beziehung zwischen Hund und PflegerIn beeinflusst werden (beispielsweise Vertrauensverhältnis vs. Unterdrückungsverhältnis) und diese kann wiederum völlig abweichen von der Art der Beziehung zwischen Hund und NeubesitzerIn – eventuell noch mehr als von einer neutralen Beurteilungsperson. Dies zeigt die Schwierigkeit bzw. Unmöglichkeit eines Wesenstests, gewisse Situationen, Beziehungen oder Umstände zu simulieren – in ihrer Diversität der Realität.

In einer Studie (DE MEESTER et al., 2008) wurde die Fähigkeit des *SAB Wesenstests* bei Hunden in Privatbesitz untersucht, Charakterzüge des Hundes bezüglich Schüchternheit/Selbstbewusstsein zu messen. Die An-/Abwesenheit der Besitzerin/ des Besitzers schien hierbei den größten Einfluss auf Änderungen der Körperhaltung der Hunde während des Wesenstest zu haben. Bei den Einzeltests in Abwesenheit der BesitzerInnen war die Körperhaltung niedriger, die die Rute wurde weiter unten gehalten, die Ohren wurden ebenfalls niedriger gehalten. DE MEESTER et al. (2011) haben diese Frage noch ein weiteres Mal aufgegriffen und die Körperhaltungen und Verhaltensstrategien von Hunden im *SAB Wesenstest* analysiert. Die Faktorenanalyse der Ergebnisse ergab, dass zwei zugrunde liegende Faktoren - Anwesenheit und Abwesenheit der BesitzerInnen - 64% der Gesamtvarianz in den Testergebnissen erklärten.

Wie bereits an mehreren Stellen vermutet, bzw. untersucht, hat die An-/Abwesenheit einer Bezugsperson erheblichen Einfluss auf das Verhalten des Hundes. Untersuchungen wie

diese bestärken die Vermutung einer mangelnden Aussagekraft von Wesenstests in Tierheimen in Abwesenheit einer Bezugsperson.

Eine „limitierte“ Version einer Bezugsperson bildet im Tierheim in der Regel das jeweilige betreuende Personal, durch das eventuell zusätzliche Informationen erlangt werden könnten. Nach den Wesenstests durch nach VAN DER BORG et al. (1991) wurde der/die jeweilige PflegerIn interviewt durch einen Fragebogen, welcher vergleichbar mit dem Test gestaltet wurde. Leider war es unmöglich, in Abgabefällen eine unbefangene Einschätzung zu erhalten, in denen die ehemaligen BesitzerInnen dem Personal über ihre Erfahrungen mit dem jeweiligen Hund berichteten. So war die Befragung der PflegerInnen in den meisten Fällen eine Mischung aus eigenen Beobachtungen und Informationen, welche durch die ehemaligen BesitzerInnen geliefert wurden. Bei Fundhunden war dieser Einfluss dementsprechend nicht vorhanden. VAN DER BORG et al. (1991) erfassten somit für ihren Wesenstest den Grad der Übereinstimmung zwischen den Personalbefragungen, den Tests und dem Auftreten des jeweiligen Verhaltens im neuen Zuhause als Referenzwert. Die Wesenstests konnten etwa 75% der später aufgetretenen Problemverhalten korrekt voraussagen, wogegen nur etwa 33% der Personalbefragungen richtig lagen. Eine Ausnahme bildete das Problem *Verunreinigung des Hauses*. Diejenigen Probleme, welche nur durch das Personal korrekt evaluiert wurden, konnten eher auf die Informationen der VorbesitzerInnen zurückgeführt werden (*Trennungsangst*, *Probleme beim Autofahren* und *Verunreinigung des Hauses*). Durch Kombination beider Evaluierungsmethoden konnten insgesamt 80% der Problemverhalten vorausgesagt werden. Auf diesem Wege konnten *Trennungsangst*, *exzessives Bellen* und das *Anspringen von Menschen* zu 100% korrekt vorausgesagt werden. Der Wesenstest konnte zwar Problemverhalten besser vorhersagen als das Personal, ergab aber auch mehr Falsch-Positive als das Personal. Die Zahl der Falsch-Negativen war dagegen bei der Befragung des Personals höher als beim Test. Mit Ausnahme der *Trennungsangst* war die Sensitivität des Wesenstests bei den Problemverhalten signifikant größer als die der Personalbefragung. Der negative prädiktive Wert des Wesenstests war signifikant höher als der der Personalbefragung bei *Ziehen an der Leine*. Die Personalbefragung erlangte lediglich gute negative prädiktive Werte in *Ungehorsam*, *Aggression gegenüber Hunden* und *Probleme beim Autofahren*, während der Wesenstest mit Ausnahme der Sensitivität für *Probleme beim Autofahren* für alle Parameter eine gute bis sehr gute Sensitivität und einen guten bis sehr guten negativen prädiktiven Wert erreichte.

Die Information über *Probleme beim Autofahren* sind scheinbar eine wertvolle Information, welche durch VorbesitzerInnen gegeben werden kann. Ungehorsam und *Aggression gegenüber Hunden* lassen sich durch das Personal gut in der täglichen Routine erfassen. Dennoch ergab der Wesenstest in diesen beiden Kategorien einen besseren negativen prädiktiven Wert. Die Information, ob der Hund die Wohnung verunreinigt, lässt sich besser aus der Befragung der/des VorbesitzerIn oder durch Beobachtung durch das Tierheimpersonal evaluieren. *Probleme beim Autofahren* und *Trennungsangst* lassen sich vom Personal dagegen kaum abschätzen.

Die „Meinung“ der NeubesitzerInnen ist meist keine ExpertInnenmeinung, und der Umgang kein „Experten-Umgang“, was eine Unterschätzung des Auftretens von Problemverhalten in der Personalbefragung erklären könnte. Außerdem könnte das Personal den Hund zu positiv porträtieren, um dessen Vermittlungschancen zu verbessern.

#### **4.2.5 Conclusio Wesenstests**

Trotz aller Bemühungen um genaue Definitionen bleiben Messungen von Verhaltensscores zu einem gewissen Grad subjektiv und können niemals so klinisch interpretierbar sein wie Unterschiede in Alter, Körpertemperatur oder Körpergewicht (PATRONEK et al., 2019). Aufgrund der Natur des Hundeverhaltens sind nur einzelne Aspekte beständig, daher können nicht alle Aspekte durch einen einzelnen Test vorausgesagt werden. Kanine Aggression ist zum Beispiel sehr komplex und kontextspezifisch und mit Sicherheit der am schwierigsten zu validierende Einzeltest, sollte dies überhaupt jemals möglich sein. Kein Wesenstest kann mit Sicherheit die Zukunft eines Hundes voraussagen (MORNEMENT et al., 2015). Die zuverlässige Vorhersage von Problemverhalten im neuen Zuhause ist unwahrscheinlich, in der Theorie und noch weniger unter der eingeschränkten Umsetzbarkeit der Inszenierung von Gegebenheiten einer realen Welt im Tierheim (PATRONEK et BRADLEY, 2016). Ebenso wenig ist ein einzelner Test im Stande, Entscheidungen über Leben und Tod bei einem Hund zu treffen (CLAY et al., 2020c). Im Vergleich zu anderen diagnostischen Tests, in denen Diagnosen übereinstimmen, reflektiert Verhalten viele verschiedene Faktoren und gleichzeitig kann ein einzelner Faktor in diverse unterschiedliche Verhaltensweisen resultieren. Verhalten wird beeinflusst durch Umweltfaktoren, frühere Erfahrungen und Erlerntes, sowie durch die medizinische Kondition.

Egal wie standardisiert ein Wesenstest ist, er kann niemals genau identisch bei zwei Gelegenheiten sein (MORNEMENT et al., 2014). Genauso wenig kann jedes mögliche Szenario oder Stimulus nach der Adoption repliziert werden. Die Umgebung im neuen Zuhause ist nicht standardisiert und von unzähligen Variablen abhängig, wie den Eigenschaften der neuen BesitzerInnen, deren Umgang mit dem Hund, dem Wissensstand und der Erfahrung, sowie der Dynamik und der Routine im neuen Zuhause. Der Vorbehalt, dass Wesenstests kein „Zuhause“ simulieren können, sollte immer im Hinterkopf behalten werden (CLAY et al., 2020c). Wesenstests können maximal einen momentanen Einblick in das Wesen des Hundes geben, vor allem bei Fundhunden, zu denen es keine Vorinformationen gibt.

PATRONEK et BRADLEY (2019), sowie PATRONEK et al. (2019) lehnen die Verwendung von Wesenstests vollkommen ab. Veranschaulicht werden deren Argumente in Abb. 3.



Abb. 3: Grafik zur Sinnhaftigkeit von Wesenstests nach PATRONEK et BRADLEY, 2016

Zu bedenken ist auch der Unterschied im Stand der Wissenschaft und im Wesenstest der Menschen, vor allem Begriffe der Statistik betreffend. PATRONEK et al. (2019) erwähnen ein Missverhältnis zwischen dem tatsächlichen Stand der Wissenschaft und dem, was die Menschen vermuten, was auf folgende Ursachen zurückgeführt werden kann:

- Verwirrung durch das Vermischen von Umgangssprache mit der wissenschaftlichen Verwendung von Wörtern wie *validiert*, *zuverlässig*, *prädiktiv* und *Übereinstimmung*
- Die Limitationen von Korrelation und Regression als statistische Methoden zum Nachweis von Übereinstimmung oder Vorhersagekraft (beispielsweise gängige Faustregeln zur Beurteilung der Übereinstimmung, wie „akzeptable Korrelation“, welche aber nicht konsistent verwendet oder universell anerkannt sind)
- Unsicherheit über den Unterschied zwischen der prädiktiven Validität eines Tests in einem wissenschaftlichen Versuch vs. der Vorhersagekraft und der Fehlerrate für individuelle Hunde in der realen Welt (beispielsweise in Bezug auf Tierheimhunde vs. Hunde in Privatbesitz und auf die entsprechende Prävalenz unerwünschten Verhaltens)
- Zusammenfassen statistischer Signifikanz und klinischer Signifikanz (beispielsweise kann die Stichprobengröße die Stärke der statistischen Signifikanz beeinflussen).

Aus den oberen Punkten resultierend ist, dass dadurch Testverfahren ein Zustand der Validität verliehen wird, obwohl Studienergebnisse dazu deutlich begrenzt sind. Die Validität eines Testverfahrens müsste beispielsweise in einer Serie von Studien bestätigt und für verschiedene Populationen angewendet werden – eine einmalige Studie ist nicht ausreichend.

Der ursprüngliche Sinn von Wesenstests – besonders geeignete Individuen für bestimmte Arbeiten zu selektieren, wird in Tierheimwesenstests aus dem Zusammenhang gerissen (PATRONEK et al., 2019). Allein die Verwendung von Wesenstests als Zusatzinformationsquelle könnte unbewusst den Glauben vermitteln, dass durch diesen in irgendeiner Art und Weise objektive voraussagende Bedeutung zuzuschreiben ist – daher ist die einmalige Verwendung im Tierheim zur Feststellung der Vermittelbarkeit allgemein in Frage zu stellen.

Die Untersuchung der Fähigkeit der Reduzierung von Rückgaberraten durch Wesenstests bedarf Vergleichsstudien mit beispielsweise Tierheimen, welche keine Wesenstests

verwenden. Als positiver Aspekt von Fehlvergaben werden wiederum die wertvollen Mehrinformationen zu Wesen und Verhalten des Hundes in einem potentiellen Zuhause erwähnt.

Die Abwesenheit vollständiger Peer-Reviews zu Wesenstests birgt die Gefahr eines pseudowissenschaftlichen Vorgehens. Replikationsfähigkeit ist jedoch eine notwendige Anforderung an wissenschaftliche Prozesse (TAYLOR et MILLS, 2006). Der Aufruf zur Durchführung weiterer Forschungen zum Thema Wesenstests im Tierheim wäre in dieser Argumentation dennoch absurd, da bereits so viele Versuche unternommen wurden in den letzten 25 Jahren und dennoch keine zufriedenstellenden Resultate erlangt wurden (PATRONEK et al., 2019). Auch ethisch sind Wesenstests unter stark künstlichen Bedingungen und unter Zeitdruck in Frage zu stellen, in Anbetracht der mangelnden wissenschaftlichen Evidenz für deren Validität und Reliabilität.

Die ASPCA® hat bereits reagiert und empfohlen, dass beispielsweise im Wesenstest gezeigte Aggression nur dann als valide Information betrachtet werden sollte, wenn sie auch in einer anderen Situation außerhalb des Wesenstests gezeigt wurde. Dennoch sollte bedacht werden, dass es sich um eine subjektive Verhaltensbeurteilung handelt und nicht als wissenschaftlich validierte Indikation für zukünftiges Verhalten interpretiert werden kann.

Ein wesentlicher Kritikpunkt ist weiters, dass die durchführenden Tierheimangestellten selten Zugang zu wissenschaftlichen Publikationen haben und Wesenstests (oft modifiziert) anwenden, ohne deren Validität kritisch betrachten zu können (MORNEMENT et al., 2010). In Anbetracht des außergewöhnlichen Umfangs dieser Diplomarbeit, ist es daher nachvollziehbar und verständlich, dass „Laien“ den Überblick schnell verlieren können aufgrund der Fülle an diversen Testsystemen. Einige Wesenstests wurden beispielsweise einfach selbst von Tierschutzorganisationen entworfen ohne Konsultation von VerhaltensexpertInnen. Weiters gibt es beispielsweise einen Wesenstest namens „*Am I Safe*“, welcher von SARAH KALNAJS, einer „zertifizierten Hundeverhaltensberaterin“ veröffentlicht wurde, zu dem es allerdings keinerlei wissenschaftliche Publikationen gibt. All diese Testprotokolle können somit nicht nach ihrer Effektivität und Validität beurteilt werden.

Falsch-positiv getestete Hunde bedeuten ein Tierschutzproblem durch die mögliche Euthanasie (MORNEMENT et al., 2010). Falsch-negativ getestete Hunde bedeuten ein Problem für das Wohlergehen der Allgemeinheit durch die mögliche Vergabe gefährlicher

Hunde. Im Sinne aller Beteiligten sollte daher die Verwendung der Beurteilungsverfahren mit äußerster Sorgfalt geschehen. Offen bleibt außerdem weiterhin die Frage der tatsächlichen Prävalenz von diversen Problemverhalten in der Tierheimpopulation.

Jede/r neue BesitzerIn hat eigene Vorstellungen und Erwartungen davon, wie das neue Familienmitglied sein sollte. Für einige Menschen ist es problematisch, wenn der Hund auf der Straße an der Leine fremde Hunde anbellt oder anknurrt (weil sie sich beispielsweise dafür schämen) – für andere Menschen ist dies völlig irrelevant (weil sie beispielsweise ländlich wohnen und kaum Hunde antreffen). Wie von den persönlichen Erwartungen menschlicher Individuen, hängt das Wesen des potentiellen neuen Hundes von den Gegebenheiten im neuen Zuhause ab, zum Beispiel ob Kleinkinder oder bereits vorhandene Haustiere akzeptiert werden müssen oder nicht. InteressentInnen wünschen sich daher beispielsweise Informationen über Umgänglichkeit des Hundes mit Menschen und anderen Tieren (CLAY et al., 2020c). Eine wahrheitsgemäße und möglichst detaillierte Beschreibung des Hundes ist Voraussetzung für eine erfolgreiche Vermittlung. Wesenstests sollten erwünschte und unerwünschte Verhaltensweisen identifizieren, um an diesen in weiterer Folge arbeiten zu können. Es ist allerdings unklar, inwiefern die Persönlichkeit und das Verhalten des Neubesitzers /der NeubesitzerIn, sowie andere Tiere und Personen im Haushalt das Verhalten des Hundes nach der Adoption beeinflussen können (CLAY et al., 2020a).

Das Monitoring der körperlichen und geistigen Gesundheit der Hunde durch das Tierheimpersonal ist existenziell, um die Verhaltensentwicklung und den Tierwohlstatus zu verfolgen. Mindestens genauso wichtig ist die fachliche Kompetenz des Tierheimpersonals, nicht nur betreffend der Verhaltensevaluierung, auch in der Wahrnehmung des Gefühlszustands des Hundes. Während einmalige Wesenstests sehr unzuverlässig sein können, könnte Beobachtungsdaten, sofern sie über die Zeit hinweg wiederholt berichtet werden, zunehmend Gültigkeit zugesprochen werden, was für die Durchführung von Ethogrammen und deren Integration in den Tierheimalltag spricht. Die neue, stressvolle, (hoffentlich) vorübergehende und sozial isolierende Tierheimumgebung beeinflusst jeden Hund, abhängig von seiner individuellen Neuroplastizität, seinen Lebenserfahrungen, seinen Genen und seinen erlernten Verhaltensweisen, unterschiedlich stark (CLAY et al., 2020c). Dazu kommt meist der Verlust der menschlichen Bezugsperson(en). Aus dieser Situation heraus Rückschlüsse auf das Verhalten in einer stabilen Umgebung nach Aufbau unterstützender

sozialer Beziehungen im neuen Zuhause zu ziehen, hat sich als sehr ineffektiv erwiesen. Es stellt sich die Frage, ob auch noch so standardisierte und deutlich verbesserte Wesenstests das komplexe Verhalten von Hunden und Problemverhalten jemals evaluieren können (CLAY et al., 2020a). Dagegen steht die jahrelange Praxis der Anwendung dieser Tests und zahlreiche Modifikationen, welche oft tierheimspezifisch sind.

Neben Verhaltensbeurteilungen können aber auch andere Maßnahmen den Vermittlungserfolg steigern (HAVERBEKE et al., 2015). Darunter fällt die Empfehlung von Trainingsstunden nach der Adoption, die Einführung von *environmental enrichment* Maßnahmen und kleinen Abänderungen der Routine im Tierheim, Stressreduktionsmaßnahmen im Tierheim, Rehabilitationsprogramme (beispielsweise gezieltes Training und Spaziergänge mit dem Hund), Präadoptionstraining (beispielsweise dem Hund anzutrainieren, bei PassantInnen in den Vorderbereich des Zwingers zu kommen und ruhig stehen zu bleiben), ausführliche Beratung der HundeinteressentInnen vor der Adoption und die Verbesserung der Präsentation des jeweiligen Hundes (beispielsweise durch Infokarten am Zwinger und Verbesserung seiner äußeren Erscheinung) oder Betreuungspatenschaften mit Option auf Fixübernahme des Hundes. Auch PATRONEK et BRADLEY (2016) sehen eine Lösung in der Maximierung der Interaktionen mit den Hunden, welche widerspiegeln sollen, was die Hunde nach der Adoption erwartet (beispielsweise Spazierengehen, Menschen kennenlernen, Spielgruppen mit anderen Hunden, Spiele, Training). In Verbindung mit einer gründlichen und objektiven Vorgeschichte (falls verfügbar) ergeben diese natürlicheren Wege der Beurteilung eine Hilfe bei der Identifizierung von Hunden, deren Verhalten von Belang sein könnte. Zusätzlich kann die Aufnahme von normalem Repertoire von Aktivitäten das Leben der Hunde im Tierheim bereichern, die nachteiligen Auswirkungen der Abgabe im Tierheim und der Einschränkungen durch das Tierheim minimieren und gleichzeitig indikativ für die typische Persönlichkeit und des Verhalten des Hundes sein – und sie somit zu besseren Kandidaten für die Adoption machen. Auch ein Basistraining für Grundgehorsam könnte während des Aufenthalts im Tierheim angelegt werden (MORNEMENT et al., 2010). Verhaltensprobleme, welche im Zuge des Tierheimalltages auftauchen und durch diesen induziert werden, sollten idealerweise direkt in ihrem Beginn erkannt werden (CLAY et al., 2020c). An dieser Stelle können Verhaltensmodifikationstrainings ansetzen. Sollten Probleme persistieren, sind Reevaluierungen und gezielte Adaptionen indiziert.



Die Reaktion auf einen Stimulus in einer Situation muss nicht mit der in einer anderen Situation übereinstimmen. Basierend auf der Studie von DÖRING (2017a) war zum Beispiel das Zeigen von verschiedenen Kommunikationssignalen (Körpersprache) der Hunde am konstantesten während wiederholter Testsituationen. Eine allgemeine Aufklärung der HundehalterInnen zur Gefahrenprävention im Alltag und zum richtigen Umgang mit dem Hund inklusive der korrekten Interpretation von Körpersignalen wäre hier ein sich erneut bestätigender Lösungsansatz. In Österreich wird dies durch den Sachkundenachweis für alle HundebesitzerInnen bereits in einigen Bundesländern umgesetzt.

Es ist wichtig zu erwähnen, dass auch durch den Hund völlig unverschuldete Faktoren zur Abgabe im Tierheim führen können, wie beispielsweise Familienprobleme oder Unterkunft. Auch bei managbaren Problemen hängt es oft eher am Engagement der BesitzerInnen als an den Fähigkeiten des Hundes. Tierheime können sich durch Schutzverträge der Haftung für Vorfälle nach der Adoption entziehen. Wird im Vergabegespräch offen über die Vorgeschichte des Hundes, dessen Verhalten während des Tierheimaufenthaltes und mögliche Bedenken der Bezugspersonen im Tierheim gesprochen, wurde meiner Meinung nach der Aufklärungspflicht nachgegangen und auf notwendiges Training oder Vorsichtsmaßnahmen (beispielsweise bei der Fütterung des Hundes) hingewiesen, was einen Haftungsausschluss nach der Vergabe rechtfertigen würde.

Eine Lösung könnte weiters der vermehrte Einsatz von Pflegestellen sein, besonders für Hunde, welche die Tierheimsituation schwer bewältigen (CLAY et al., 2020a; MORNEMENT et al., 2015). Weitere Studien wären sinnvoll, um zu untersuchen, ob Pflegestellen eine höhere Voraussagekraft der unterschiedlichen Messinstrumente (Fragebogen, Wesenstest, Ethogramm) erreichen. Eine Kombination aus Informationen von Fragebögen von abgebenden TierbesitzerInnen, gemeinsam mit der Durchführung von Ethogrammen durch das Tierheimpersonal (über den Akklimatisierungszeitraum hinaus) dürfte meiner Ansicht nach mit der Unterstützung durch Pflegestellen und/oder Patenschaften, welche gezielte tägliche Interaktionen bereitstellen, das geeignetste Portfolio sein, um den Charakter und das Wesen – und dadurch die Vermittelbarkeit in die ideale neue Familie zu gewährleisten.

Meiner Meinung nach wiegt der Nutzen eines Wesenstests den damit verbundenen Aufwand nicht auf und schadet dem getesteten Individuum eher, je nach Konsequenz aus den gewonnenen Ergebnissen, als ihm zu nutzen. Sowohl ein Bestehen, als auch ein

Nichtbestehen bricht die Komplexität des Verhaltens und der Persönlichkeit eines hochentwickelten Lebewesens auf ein schlichtes Testergebnis herab und gibt vor, aus diesem Ergebnis heraus gerechtfertigte Schlüsse zu erlauben. Testergebnisse sollten daher mit Vorbehalt betrachtet werden. Der Druck und die Bemühungen, Tierheimhunde beurteilen, kategorisieren und wahrheitsgemäß beschreiben zu wollen, ist absolut nachvollziehbar, doch ich denke, dass eine genaue Vorhersage für Verhaltensentwicklung ebenso wenig möglich ist, wie beispielsweise für Krankheitsverläufe. Ein klinischer diagnostischer Test erlaubt in der Regel nur eine Aussage über den Status Quo (beispielsweise *Sind Mikrofilarien im heute abgenommenen Blut vorhanden?*) und keine Voraussage über die Zukunft. Die Aussage im Wesenstests wäre demnach zum Beispiel: *Der Hund hat in einem bestimmten Einzeltest an einem bestimmten Tag unter gewissen Umständen eine gewisse Reaktion gezeigt.*

### **4.3 Ethogramme**

Zu Ethogrammen gibt es lediglich drei Publikationen und keines dieser Ethogramme wurde wiederholt untersucht. Auch haben alle drei Studien nur die Evidenzklasse *Bronze* erreicht. Die Ergebnisse sind weiters kaum vergleichbar, da die untersuchten Messparameter und die diagnostischen Testkriterien stark variieren.

#### **4.3.1 Objektivität von Ethogrammen**

Trotz des Zeitaufwandes könnten Ethogramme in anderen Aspekten der Objektivität durchaus Vorteile mit sich bringen. So finden beispielsweise bei CLAY et al. (2019) keinerlei Interaktionen mit dem zu testenden Hund statt, welche es zu standardisieren gilt. Dagegen steht allerdings die Tatsache, dass viele Verhaltensweisen daher nicht gezeigt werden (können). In den anderen beiden Studien dagegen wurde interagiert – standardisierte Spaziergänge/ Wesenstests in DE PALMA et al. (2005) und die Erfassung sämtlicher Interaktionen mit dem Hund in GOOLD et NEWBERRY (2017). Letzteres fand zwar nicht standardisiert statt, wäre allerdings weniger zeitaufwändig (und daher praktikabler) und würde realistische Situationen eventuell besser widerspiegeln.

#### **4.3.2 Reliabilität von Ethogrammen**

Wiederholungstests können nur bei standardisierten Abläufen stattfinden, was in CLAY et al. (2019) und GOOLD et NEWBERRY (2017) nicht der Fall war. DE PALMA et al. (2005) hätten

die Möglichkeit gehabt, da sie die meisten Interaktionen mehrmals durchgeführt haben (z.B. zwei Spaziergänge zu jeweils 30 Minuten, haben allerdings ihren Fokus anderweitig gesetzt.

Ethogramme erlauben dagegen aber Einblicke in die Verhaltensentwicklung von Hunden. Über den Beobachtungszeitraum gab es beispielsweise eine signifikante Abnahme von angstbezogenen Verhaltensweisen, so war in den ersten fünf Tagen eine Verminderung von Angstindikatoren wie das Einziehen der Rute zu beobachten (CLAY et al., 2019).

In den spärlichen Untersuchungen der Inter-Rater-Reliabilität ist diese jeweils recht hoch gewesen (0.77 in GOOLD et NEWBERRY, 2017 und 0.89 in DE PALMA et al., 2005).

#### **4.3.3 Validität von Ethogrammen**

Leider wurde in keiner der Studien das Verhalten nach der Adoption untersucht und mit den Ergebnissen der Ethogramme verglichen, eine Aussage in diesem Aspekt ist also nicht möglich.

#### **4.3.4 Conclusio Ethogramme**

Die Beobachtungen finden vor allem in der Tierheimroutine unter alltäglichen Bedingungen statt und nicht im Rahmen einer Testsituation. Hierfür erstellen mehrere BeobachterInnen ein Profil des Hundes über einen längeren Zeitraum hinweg. Ethogramme könnten, sofern sie standardisiert sind, eine objektivere und zuverlässigere Alternative zu Fragebögen und Wesenstests darstellen. Diese beiden Kriterien (Objektivität, Reliabilität) haben scheinbar bereits ein höheres Potential, zur Validität fehlen allerdings jegliche Messwerte, hier besteht Forschungsbedarf. Zwar ist der Zeitaufwand für die Auswertung der einzelnen Beobachtungen höher, dafür entfällt die aufwendige Organisation eines Wesenstests.

Im humanen Sozialbereich sind Einträge und Aufzeichnungen zu Beobachtungen und Interaktionen mit den KlientInnen durch die Betreuenden bereits fester Bestandteil des Arbeitsalltages, eventuell könnte eine ähnliche Vorgehensweise wie im Ethogramm nach GOOLD et NEWBERRY (2017) in den Tierheimalltag integriert werden. Hierbei wurden sämtliche Interaktionen des Hundes mit fremden Personen beobachtet und dokumentiert. Denkbar wären daher auch im Tierheimalltag kurze tagebuchähnliche Einträge durch die Mitarbeitenden am Ende der Schicht zu den am jeweiligen Tag betreuten Hunden, beispielsweise bezüglich Verhaltensweisen während eines Spazierganges, eines Aufeinandertreffens

mit InteressentInnen oder über Auffälligkeiten in der täglichen Routine, wie zum Beispiel beim Füttern.

## 5 Zusammenfassung

Unzählige Tierheime und Tierschutzorganisationen weltweit sehen eine ihrer Aufgaben darin, das Verhalten und die Vermittelbarkeit der Hunde in ihrer Obhut zu beurteilen. Dies soll in erster Linie zu einem besseren Zusammenpassen mit potentiellen zukünftigen BesitzerInnen führen und jene Hunde identifizieren, welche aufgrund von Verhaltensproblemen (noch) nicht für eine Vermittlung geeignet sind.

Ziel dieser Arbeit sollte es sein, diese vielfältigen Beurteilungsverfahren auf ihre Praxis-tauglichkeit und ihre Verlässlichkeit zu untersuchen, indem vorhandene Veröffentlichungen auf die Erfüllung wichtiger diagnostischer Testparameter evaluiert wurden.

Mithilfe verschiedener Suchmaschinen wie Pubmed, Science Direkt oder Scopus wurden 97 Studien zum Thema *Wesensbeurteilung im Tierheim* gefunden und unter Anwendung des PRISMA Statements (Preferred Reporting Items for Systematic Review and Meta-Analyses) weiter selektiert, sodass letztendlich 34 Publikationen in die systematische Literaturübersicht eingeschlossen werden konnten. In weiterer Folge wurden diese Studien in Evidenzklassen eingeteilt (modifiziert nach SANTESSO et al., 2006). Keine der Publikationen konnte die Kriterien zur Einteilung in Evidenzklasse Platin oder Gold erfüllen. Lediglich 4 der 34 Studien konnten der Evidenzklasse Silber zugeordnet werden, alle weiteren Studien nur Bronze.

Anschließend wurden die Studien zu Verhaltensevaluierungsmethoden, welche in Fragebögen (Verhaltensbeurteilung durch VorbesitzerInnen), Wesenstests und Ethogramme eingeteilt werden konnten, auf die Untersuchung der drei diagnostischen Testparameter Objektivität, Zuverlässigkeit (Reliabilität) und Gültigkeit (Validität) betrachtet. Vorhandene Ergebnisse wurden ausgewertet und verglichen.

Leider existieren zu keiner Methode der Verhaltensevaluierung ausreichende Untersuchungen hinsichtlich aller relevanter diagnostischer Testparameter.

Die Ergebnisse gemessener diagnostischer Testparameter variierten stark in Bezug auf den jeweiligen Messparameter. Besondere Bedeutung hat im Zusammenhang mit Verhaltensbeurteilungen die prognostische Validität, welche messen soll, inwiefern eine Methode das Verhalten nach der Adoption voraussagen kann. Ein Vergleich der Methoden miteinander war aufgrund der starken Varianz an Testbedingungen, der verwendeten Messparameter und

statistischen Messwerte nur eingeschränkt möglich. Keines der publizierten Evaluierungsverfahren stellte sich als ausreichend objektiv, reliabel und valide heraus.

## 6 Summary

Many animal shelters and animal welfare organizations worldwide are in need to assess the behaviour and adoptability of the dogs in their care. This should primarily lead to a better match with potential future owners and identify those dogs that are not (yet) suitable for placement due to behavioural problems.

The aim of this work is to examine these diverse assessments for their practicability and their reliability by evaluating existing publications for the fulfillment of important diagnostic test parameters.

With the help of various search engines such as Pubmed, Science Direct or Scopus, 97 studies were found dealing with behaviour assessment in animal shelters and further selected using the PRISMA statement (Preferred Reporting Items for Systematic Review and Meta-Analyses). Finally, 34 publications were included in this systematic literature review. These studies were then divided into evidence classes (modified from SANTESSO et al., 2006). None of the publications could meet the criteria for classification in evidence class *platinum* or *gold*. Only 4 of 34 studies could be assigned to the evidence class *silver*, all other studies only to *bronze*.

Then the studies on behaviour evaluation methods, were examined for the three diagnostic test parameters objectivity, reliability and validity, classified on into three subcategories namely questionnaires (behavioural assessment by relinquishing owners), behaviour assessments and ethograms. Existing results were extensively evaluated and compared.

In conclusion no adequate method of behavioural evaluation in shelter dogs with regard to all relevant diagnostic test parameters was found.

The results of measured diagnostic test parameters varied greatly with respect to the particular measurement value in question. In connection with behavioural assessments, the prognostic validity, which is intended to measure the extent to which a method can predict behaviour after adoption, is of particular importance.

A comparison of the methods with one another was only possible to a very limited extent due to the great variance in test conditions, the measurement parameters used and the statistical measurement values published.

None of the evaluation procedures turned out to be sufficiently objective, reliable and valid.



## 7 Begriffserklärungen

Wesenstest – Der Begriff „Wesenstest“ wird hier zusammenfassend für sämtliche Beurteilungsverfahren von Hundeverhalten in der Literatur verwendet, inkludierend “temperament test“, “behavioral assessment“, “behavior evaluation“, “personality test“, “behavioral test“, „Aggressionstest“, „Verhaltenstest“.

Tierheim – Der Begriff „Tierheim“ wird hier stellvertretend für sämtliche Behausungen verwendet, in denen Hunde aus Gründen der Weitervermittlung verwahrt werden, inkludierend “rescue shelters“, städtische Tierheime, private Tierheime, vermittelnde Tierschutzvereine.

Hund – Der Begriff „Hund“ wird in dieser Arbeit geschlechterübergreifend für Rüden und Hündinnen verwendet. Ebendies gilt für den Begriff „Artgenosse“.

## 8 Literaturverzeichnis

2007. Behaviour assessment tool may help shelter adoptions. *Journal of the American Veterinary Medical Association*, 231(1):22.
2020. American Society for the Prevention of Cruelty to Animals <https://aspcapro.org> (Zugriff 15.09.2020)
2021. American Society for the Prevention of Cruelty to Animals. <https://aspca.org> (Zugriff 12.04.2021)
- Alberghina D, Piccione G, Pumilia G, Gioé M, Rizzo M, Raffo P, Panzera M. 2019. Daily fluctuation of urine serotonin and cortisol in healthy shelter dogs and influence of intraspecific social exposure. *Physiology & Behavior*, 206(2019):1-6.
- Arena L, Wemelsfelder F, Messori S, Ferri N, Barnard S. 2019. Development of a fixed list of terms for the Qualitative Behavioural Assessment of shelter dogs. <https://doi.org/10.1371/journal.pone.0212652> (Zugriff 09.09.2020).
- Arena L, Wemelsfelder F, Messori S, Ferri N, Barnard S. 2017. Application of Free Choice Profiling to assess the emotional state of dogs housed in shelter environments. *Applied Animal Behaviour Science*, 195(2017):72-79.
- Arena L, Berteselli GV, Lombardo F, Candeloro L, Dalla Villa P, De Massis F. 2019. Application of a welfare assessment tool (Shelter Quality Protocol) in 64 Italian long-term dogs' shelters: welfare hazard analysis. *Animal Welfare*, 28(3):353-363.
- Arvelius P, Blixt C, Svartberg K, Nyberg ST. 2012. A new behavior and personality test to be used as a tool in dog breeding. *Journal of Veterinary Behavior*, Vol.7(1):57.
- Arruda EC, Garcia RCM, Oliveira ST. 2020. The welfare of dogs from municipal shelters in the state of Paraná, Brazil, under the Shelter Quality protocol. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 72(2):346-354.
- Barnard S, Pedernera C, Velarde A, Dalla Villa P. 2014. Shelter Quality. Welfare Assessment Protocol for Shelter Dogs. <http://www.carodog.eu/wp-content/uploads/2014/02/Shelter-Quality-Protocol-2014.pdf> (Zugriff 09.09.2020)

- Barnard S, Siracusa C, Reisner I, Valsecchi P, Serpell JA. 2012. Validity of model devices used to assess canine temperament in behavioral tests. *Applied Animal Behaviour Science*, 138(1-2):79-87.
- Barnard S, Pedernera C, Candeloro L, Ferri N, Velarde A, Dalla Villa P. 2015. Development of a new welfare assessment protocol for practical application in long-term dog shelters. *The Veterinary record*, 178(1):18.
- Barnard S, Kennedy D, Watson R, Valsecchi P, Arnott G. 2019. Revisiting a Previously Validated Temperament Test in Shelter Dogs, Including an Examination of the Use of Fake Model Dogs to Assess Conspecific Sociability. *Animals (Basel)*, 9(10):835.
- Barrera G, Fagnani J, Mustaca AE, Bentosela M, Giamal Y. 2013. Evaluation of the Temperament of shelter dogs and family pets. A comparative study. *Revista Colombiana de Psicología* 22(2):307-320.
- Bauer AE, Jordan M, Colon M, Shreyer T, Croney CC. 2017. Evaluating FIDO: Developing and pilot testing the Field Instantaneous Dog Observation tool. *Pet Behaviour Science*, 4(2017):1-15.
- Beerda B, Schilder MBH, Van Hooff JARAM, De Vries HW, Mol JA. 1999. Chronic Stress in Dogs Subjected to Social and Spatial Restriction. I. Behavioral Responses. *Physiology & Behavior*, 66(2):233-242.
- Beerda B, Schilder MBH, Bernadina W, Van Hooff JARAM, De Vries HW, Mol JA. 1999. Chronic Stress in Dogs Subjected to Social and Spatial Restriction. II. Hormonal and Immunological Responses. *Physiology & Behavior*, 66(2):243-254.
- Bennett SL. 2010. Temperament Tests: What we do and do not know about them [Vortrag]. <https://www.maddiesfund.org/assets/documents/Resource%20Library/Temperament%20Testing%20Presentation%20Notes.pdf> (Zugriff 08.09.2020).
- Bennett SL, Weng HY, Walker SL, Placer M, Litster A. 2015. Comparison of SAFER™ behavior assessment results in shelter dogs at intake and after a 3-day acclimation period. *Journal of applied animal welfare science*, 18(2):153-168.

- Bennett SL, Litster A, Weng HY, Walker SL, Luescher AU. 2012. Investigating behavior assessment instruments to predict aggression in dogs. *Applied Animal Behaviour Science*, 141(3-4):139-148.
- Berteselli GV, Arena L, Candeloro L, Dalla Villa P, De Massis F. 2019. Interobserver agreement and sensitivity to climatic conditions in sheltered dogs' welfare evaluation performed with welfare assessment protocol (Shelter Quality protocol). *Journal of Veterinary Behavior*, 29(1-2):45-52.
- Binder R. 2018. Euthanasie von Heimtieren: Das Tierschutzrecht zwischen Lebensschutz und Leidverkürzung. *Wiener Tierärztliche Monatsschrift*, 105(2018):119-128.
- Bollen KS, Horowitz J. 2008. Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Applied Animal Behaviour Science*, 112(1-2):120-135.
- Brady K, Cracknell N, Zulch H, Mills DS. 2018. A Systematic Review of the Reliability and Validity of Behavioural Tests Used to Assess Behavioural Characteristics Important in Working Dogs. *Frontiers in Veterinary Science*, 5(2018):103.
- Bräm M, Doherr MG, Lehmann D, Mills D, Steiger A. 2008. Evaluating aggressive behavior in dogs: a comparison of 3 tests. *Journal of Veterinary Behavior*, 3(4):152-160.
- Brown KM, Feuerbacher EN, Hall NJ, Protopopova A. 2020. Minor procedural variations affect canine behavior during sociability assessments. *Behavioural Processes*, 177(2020):104145.
- Buser K. 2007. *Kurzlehrbuch Medizinische Psychologie – Medizinische Soziologie*. München, Deutschland: Urban & Fisher Verlag/ Elsevier GmbH.
- Cafazzo S, Maragliano L, Bonanni R, Scholl F, Guarducci M, Scarcella R, Di Paolo M, Pontier D, Lai O, Carlevaro F, Bucci E, Cerini N, Carlevaro L, Alfieri L, Fabtini C, Natoli E. 2014. Behavioural and physiological indicators of shelter dogs' welfare: Reflections on the no-kill policy on free-ranging dogs in Italy revisited on the basis of 15 years of implementation. *Physiology & Behavior*, 133(2014):223-229.

- Canejo-Teixeira R, Almiro PA, Serpell JA, Baptista LV, Niza MMRE. 2018. Evaluation of the factor structure of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) in European Portuguese. <https://doi.org/10.1371/journal.pone.0209852> (Zugriff 08.09.2020).
- Christensen E, Scarlett J, Campagna M, Houpt K. 2007. Aggressive Behavior in Adopted Dogs (*Canis Familiaris*) that Passed a Temperament Test. *Applied Animal Behaviour Science*, 106(2007):85-95.
- Clay L, Paterson MBA, Bennett PC, Perry G, Phillips CCJ. 2020. Comparison of Canine Behaviour Scored Using a Shelter Behaviour Assessment and an Owner Completed Questionnaire, C-BARQ. *Animals (Basel)*, 10:1797. (Clay et al., 2020b)
- Clay L, Paterson MBA, Bennett PC, Perry G, Phillips CCJ. 2020. Do Behaviour Assessments in a Shelter Predict the Behaviour of Dogs Post-Adoption?. *Animals (Basel)*, 10(7):1225. (Clay et al., 2020a)
- Clay L, Paterson MBA, Bennett PC, Perry G, Phillips CCJ. 2019. Early Recognition of Behaviour Problems in Shelter Dogs by Monitoring them in their Kennels after Admission to a Shelter. *Animals (Basel)*, 9(11):875.
- Clay L, Paterson M, Bennett P, Perry G, Rohlf V, Phillips CJC. 2020. In defense of canine behavioral assessments in shelters: Outlining their positive applications. *Journal of Veterinary Behavior*, 38:74-81. (Clay et al., 2020c)
- Cohen SE, Todd PM. 2019. Stated and revealed preferences in companion animal choice. *Behavior Research Methods*, 51(2019):1498-1509.
- Coren S. 2004. *How Dogs Think – Understanding the Canine Mind*. New York, USA: Free Press.
- D'Arpino, Dowling-Guyer S, Shabelansky A, Marder AR, Patronek GJ. 2012. The Use and Perception of Canine Behavioral Assessments in Sheltering Organizations [Vortrag]. In: 2012 ACVB/AVSAB Veterinary Behavior Symposium, 27-30.

- Dalla Villa P, Barnard S, De Fede E, Podaliri M, Candeloro L, Di Nardo A, Siracusa C, Serpell JA. 2013. Behavioural and physiological responses of shelter dogs to long-term confinement. *Veterinaria italiana*, 49(2):231-241.
- Dalla Villa P, Barnard S, Di Nardo A, Iannetti L, Vulpiani MP, Trentini R, Serpell JA, Siracusa C. 2017. Validation of the Socially Acceptable Behaviour (SAB) test in a Central-Italy pet dog population. *Veterinaria Italiana* 2017, 53(1):61-70.
- De Palma C, Viggiano E, Barillari E, Palme R, Dufour AB, Fantini C, Natoli E. 2005. Evaluating the Temperament in Shelter Dogs. *Behaviour*, 142(9):1313-1334.
- De Meester RH, Pluijmakers J, Vermeire S, Laevens H. 2011. The use of the socially acceptable behavior test in the study of temperament of dogs. *Journal of Veterinary Behavior*, 6(4):211-224.
- De Meester RH, De Bacquer D, Peremans K, Vermeire S, Planta DJ, Coopman F, Audenaert K. 2008. A preliminary study on the use of the Socially Acceptable Behavior test as a test for shyness/confidence in the temperament of dogs. *Journal of Veterinary Behavior*: 3(4):161-170.
- Diederich C, Giffroy JM. 2006. Behavioural testing in dogs: A review of methodology in search for standardisation. *Applied Animal Behaviour Science*, 97:51-72.
- Diesel G, Brodbelt D, Pfeiffer DU. 2008. Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. *Applied Animal Behaviour Science*, 115(3-4):171-181.
- Döring D, Haberland BE, Ossig A, Küchenhoff H, Dobenecker B, Hack R, Schmidt J, Erhard MH. 2016. Behavior of laboratory beagles: Assessment in a standardized behavior test using novel stimuli and situations. *Journal of Veterinary Behavior*, 11:18-25.
- Döring D, Haberland BE, Bauer A, Dobenecker B, Hack R, Schmidt J, Erhard MH. 2017. Consistency in behavior: Evaluation of behavior tests in laboratory beagles. *Journal of Veterinary Behavior*, 21:59-63. (DÖRING et al., 2017a)

- Döring D, Nick O, Bauer A, Küchenhoff H, Erhard MH. 2017. Behavior of Laboratory Dogs Before and After Rehoming in Private Homes. *ALTEX Alternatives to Animal Experimentation*, 34(1):133-147. (DÖRING et al., 2017b)
- Dowling-Guyer S, Marder A, D'Arpino S. 2011. Behavioral traits detected in shelter dogs by a behavior evaluation. *Applied Animal Behaviour Science*, 130(3):107-114.
- Duffy DL, Kruger KA, Serpell JA. 2014. Evaluation of a behavioral assessment tool for dogs relinquished to shelters. *Preventive veterinary medicine*, 117(3-4):601-609.
- Duffy DL, Serpell JA. 2012. Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science*, 138(2012):99-109.
- Ettlin D, Gallo LM. 2019. *Das Kiefergelenk in Funktion und Dysfunktion*. Stuttgart, Deutschland: Verlagsgruppe Georg Thieme.
- Fernandes JG, Olsson IAS, DeCastro ACV. 2017. Do aversive-based training methods actually compromise dog welfare?: A literature review. *Applied Animal Behaviour Science* 196(2017):1-12.
- Feuerbacher EN, Wynne CDL. 2012. Relative Efficacy of Human Social Interaction and Food as Reinforcers for Domestic Dogs and Hand-reared Wolves. *Journal of the Experimental Analysis of Behavior*, 98(1):105-129.
- Frank C. 2007. *Entwicklung und Anwendung eines Verhaltenstests für Tierheimhunde sowie die Untersuchung der Verhaltensentwicklung im Tierheim [Dissertation]*. Berlin: Freie Universität Berlin.
- González-Ramírez MT, Quezada-Berumen L, Landero-Hernández R. 2017. Assessment of canine behaviors using C-BARQ in a sample from Northern Mexico. *Journal of Veterinary Behavior*, 20(7-8):52-58.
- Goold C, Newberry RC. 2017. Modelling personality, plasticity and predictability in shelter dogs. *Royal Society Open Science*, 4(9):170618. DOI 10.1098/rsos.170618 (Zugriff 08.09.2020).

- Gras JC. 2016. Behavioral test for the dog's temperament evaluation and their adaption in the domestic environment. <https://core.ac.uk/download/pdf/132086194.pdf> (Zugriff 08.09.2020).
- Gunter LM, Barber RT, Wynne CDL. 2018. A canine identity crisis: Genetic breed heritage testing of shelter dogs. <https://doi.org/10.1371/journal.pone.0202633> (Zugriff 09.09.2020).
- Gunter L. 2018. Understanding the Impacts of Breed identity, Post-Adoption and Fostering Interventions, & Behavioral Welfare of Shelter Dogs [Dissertation]. Tempe: Arizona State University.
- Hare E, Kelsey KM, Serpell JA, Otto CM. 2018. Behavior Differences Between Search-and-Rescue and Pet Dogs. *Frontiers in Veterinary Science*, 5:118.
- Haverbeke A, Pluijmakers J, Diederich C. 2015. Behavioral evaluations of shelter dogs: Literature review, perspectives, and follow-up within the European member states's legislation with emphasis on the Belgian situation. *Journal of Veterinary Behavior*, 10(1):5-11.
- Hennessy MB, Voith VL, Mazzei SJ, Buttram J, Miller DD, Linden F. 2001. Behavior and cortisol levels of dogs in a public animal shelter, and exploration of the ability of these measures to predict problem behavior after adoption. *Applied Animal Behaviour Science*, 73(3):217-233.
- Hewson CJ, Hiby E, Bradshaw JWS. 2007. Assessing quality of life in companion and kennelled dogs: A critical review. *Animal welfare*, 16:89-95.
- Hiby EF, Rooney NJ, Bradshaw JWS. 2006. Behavioural and physiological responses of dogs entering re-homing kennels. *Physiology & Behavior*, 89(3):385-391.
- Hsu Y, Serpell JA. 2003. Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association*, 223(9):1293-1300.



- Iversen S. 2008. Ein Aggressionstest bei Berliner Stadthunden - Feststellung vorkommender Aggressionsformen, Reizschwellen und Einflussfaktoren für aggressives Verhalten [Dissertation]. Berlin: Freie Universität Berlin.
- Jones AC. 2008. Development and validation of a dog personality questionnaire [Dissertation]. Austin: The University of Texas at Austin.
- Jones AC, Gosling SD. 2005. Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science*, 95(2005):1-53.
- Josa JM, Suarez P. 2008. Initial assessment of canine behaviour in an animal shelter. *Revista oficial de la Asociación Veterinaria Española de Especialistas en Pequeños Animales*, 28(3):173-177.
- Kiddie JL, Collins LM. 2014. Development and validation of a quality of life assessment tool for use in kennelled dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, 158(2014):57-68.
- Kiddie J, Collins L. 2015. Identifying environmental and management factors that may be associated with the quality of life of kennelled dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, 167(2015):43-55.
- Kis A, Klausz B, Persa E, Miklósi À, Gácsi M. 2014. Timing and presence of an attachment person affect sensitivity of aggression tests in shelter dogs. *The Veterinary record*, 174(8):196.
- Klausz B, Kis A, Persa E, Gácsi M. 2009. Human-directed aggression in shelter dogs: how to test for better prediction of outcomes. *Journal of Veterinary Behavior*, 4(2):78.
- Klausz B, Kis A, Persa E, Miklósi À, Gácsi M. 2013. A Quick Assessment Tool for Human-Directed Aggression in Pet Dogs. *Aggressive Behavior*, Vol. 9999(2013):1-11.
- Kogan LR, Hellyer PW, Rishniw M, Schoenfeld-Tacher R. 2020. Veterinary Behavior: Assessment of Veterinarians' Training, Experience, and Comfort Level with Cases. *Journal of veterinary medical education*, 47(2):158-169.

- Kruger KA, Duffy DL, Serpell JA. 2009. Evaluation of a new screening tool for assessing temperament and behavior in dogs relinquished to shelters [Konferenzbeitrag]. ISAZ 18<sup>th</sup> Annual Conference, Kansas City (Montana, USA), Seite 97. <http://www.isaz.net/conferences/ISAZ2009.pdf> (Zugriff 08.09.2020).
- Lambert K, Coe J, Niel L, Dewey C, Sargeant JM. 2015. A systematic review and meta-analysis of the proportion of dogs surrendered for dog-related and owner-related reasons. *Preventive Veterinary Medicine*, 118(1):148-160.
- Lexikon der Biologie, Spektrum Akademischer Verlag Heidelberg  
<https://www.spektrum.de/lexikon/biologie> (Zugriff 15.09.2020)
- Ley JM, Bennett PC, Coleman GJ. 2009b. A refinement and validation of the Monash Canine Personality Questionnaire. *Applied Animal Behaviour Science*, 116:220-227.
- Ley JM, McGreevy P, Bennett PC. 2009a. Inter-rater and test-retest reliability of the Monash Canine Personality Questionnaire-Revised (MCPQ-R). *Applied Animal Behaviour Science*, 119:85-90.
- Loyer CJ. 2019. Social Behavior and Cognitive Bias in the Domestic Dog (*Canis Familiaris*) [Dissertation]. Washington: University of Washington.
- Lucidi P, Bernabó N, Panunzi M, Dalla Villa P, Mattioli M. 2005. Ethotest: A new model to identify (shelter) dogs' skills as service animals or adoptable pets. *Applied Animal Behaviour Science*, 95(1-2):103-122.
- Marder AR, Shabelansky A, Patronek GJ, Dowling-Guyer S, Segurson D'Arpino S. 2013. Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption. *Applied Animal Behaviour Science*, 148(1-2):150-156.
- Marder A, Voith V. 2010. The American Shelter Dog: Identification of dogs by personality. *Journal of Veterinary Behavior Clinical Applications and Research*, 5(1):26-26.
- Marder A, Dowling-Guyer S. 2010. Shelter dog behavioral assessments: Behavioral coding and personality. *Journal of Veterinary Behavior*, 5(1):27.

- Matters S. 2016. Intra-rater and Inter-rater Reliability of the ASPCA's Behavior Evaluation of Fearful Dogs [Dissertation]. New York: City University of New York.
- McCobb E, Dowling-Guyer S. 2016. Welfare assessments for long-term housing in animal shelters [Editorial]. *The Veterinary record*, 178(1):16-17.
- McGuire B, Orantes D, Xue S, Parry S. 2020. Abilities of Canine Shelter Behavioral Evaluations and Owner Surrender Profiles to Predict Resource Guarding in Adoptive Homes. *Animals (Basel)*, 10(9):E1702
- McGuire B. 2019. Characteristics and Adoption Success of Shelter Dogs Assessed as Resource Guardians. *Animals (Basel)*, 9(11):982.
- Medistat GmbH. 2021. <https://www.medistat.de/glossar/korrelation-assoziation/rangkorrelationskoeffizient-nach-spearman/> (Zugriff 16.04.2021)
- Menchetti L, Righi C, Guelfi G, Enas C, Moscati L, Mancini S, Diverio S. 2019. Multi-Operator Qualitative Behavioural Assessment for dogs entering the shelter. *Applied Animal Behaviour Science*, 213(2019):107-116.
- Mendl M, Burman OHP, Parker RMA, Paul ES. 2009. Cognitive bias as an inductor of animal emotion and welfare: emerging evidence and underlying mechanisms. *Applied Animal Behaviour Science*, 118:161-181.
- Meyer I, Forkman B. 2014. Dog and owner characteristics affecting the dog-owner relationship. *Journal of Veterinary Behavior*, 9(2014):143-150.
- Michelazzi M, Levi D, Fossati P, Scaglia E. 2010. Evaluation of sheltered dangerous dogs. *Journal of Veterinary Behavior*, 5(1):35.
- Miller KA, Touroo R, Spain CV, Jones K, Reid P, Lockwood R. 2016. Relationship Between Scarring and Dog Aggression in Pit Bull-Type Dogs Involved in Organized Dogfighting. *Animals (Basel)*, 6(11):72.
- Mirkó E, Kubinyi E, Gácsi M, Miklósi Á. 2012. Preliminary analysis of an adjective-based dog personality questionnaire developed to measure some aspects of personality in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, 138(2012):88-98.

- Mirkó E, Dóka A, Miklósi Á. 2013. Association between subjective rating and behaviour coding and the role of experience in making video assessments on the personality of the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, 149(1-4):45-54.
- Mohan-Gibbons H, Weiss E, Slater M. 2012. Preliminary Investigation of Food Guarding Behavior in Shelter Dogs in the United States. *Animals (Basel)*, 2(3):331-346.
- Mohan-Gibbons H, Dolan ED, Reid P, Slater MR, Mulligan H, Weiss E. 2018. The Impact of Excluding Food Guarding from a Standardized Behavioral Canine Assessment in Animal Shelters. *Animals (Basel)*, 8(2):27.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PloS Med*, 6(7):e1000097. <https://doi.org/10.1371/journal.pmed.1000097> (Zugriff 04.10.2020).
- Mornement K, Toukhsati S, Coleman G, Bennett P. 2009. Reliability, validity and feasibility of existing tests of canine behavior [Vortrag]. (Zugriff 08.09.2020).
- Mornement KM, Coleman GJ, Toukhsati SR, Bennett PC. 2015. Evaluation of the predictive validity of the Behavioural Assessment for Re-homing K9's (B.A.R.K.) protocol and owner satisfaction with adopted dogs. *Applied Animal Behaviour Science*, 167:35-42.
- Mornement KM, Coleman GJ, Toukhsati S, Bennett PC. 2014. Development of the behavioural assessment for re-homing K9's (B.A.R.K.) protocol. *Applied Animal Behaviour Science*, 151(2):75-83.
- Mornement KM, Coleman GJ, Toukhsati S, Bennett PC. 2010. A Review of Behavioral Assessment Protocols Used by Australian Animal Shelters to Determine the Adoption Suitability of Dogs. *Journal of applied animal welfare science*, 13(4):314-329.
- Mornement KM, Coleman GJ, Toukhsati S, Bennett PC. 2009. Assessing shelter dogs for adoption suitability – an Australian perspective. *Journal of Veterinary Behavior*, 4(2):85-86.

- Mugenda L, Shreyer T, Cronney C. 2019. Refining canine welfare assessment in kennels: Evaluating the reliability of Field Instantaneous Dogs Observation (FIDO) scoring. *Applied Animal Behaviour Science*, 221(2019):104874.
- Nagasawa M, Tsujimura A, Tateishi K, Mogi K, Ohta M, Serpell JA, Kikusui T. 2011. Assessment of the factorial structures of the C-BARQ in Japan. *The Journal of Veterinary Medical Science*, 73(7):869-875.
- Netto WJ, Planta DJU. 1997. Behavioural testing for aggression in the domestic dog. *Applied Animal Behaviour Science*, 52(3-4):243-263.
- Normando S, Di Raimondo G, Bellaio E. 2019. An investigation using different data gathering methods into the prevalence of behavioral problems in shelter dogs - A pilot study. *Journal of Veterinary Behavior*, 30(3-4):1-8.
- Owczarczak-Garstecka SC, Burman OHP. 2016. Can Sleep and Resting Behaviours Be Used as Indicators of Welfare in Shelter Dogs (*Canis lupus familiaris*)?. <https://doi.org/10.1371/journal.pone.0163620> (Zugriff 09.09.2020).
- Passantino A, Quartarone V, Pediliggeri MC, Rizzo M, Piccione G. 2014. Possible application of oxidative stress parameters for the evaluation of animal welfare in sheltered dogs subjected to different environmental and health conditions. *Journal of Veterinary Behavior*, 9(6):290-294.
- Patronek GJ, Bradley J. 2016. No better than flipping a coin: Reconsidering canine behavior evaluations in animal shelters. *Journal of Veterinary Behavior*, 15(2016):66-77.
- Patronek GJ, Bradley J, Arps E. 2019. What is the evidence for reliability and validity of behavior evaluations for shelter dogs? A prequel to "No better than flipping a coin". *Journal of Veterinary Behavior*, 31(2019):43-58.
- Petak I. 2013. Communication patterns within a group of shelter dogs and implications for their welfare. *Journal of applied animal welfare science*, 16(2):118-139.
- Planta JUD, De Meester RHWM. 2007. Validity of the Socially Acceptable Behavior (SAB) test as a measure of aggression in dogs towards non-familiar humans. *Vlaams Diergeneeskundig Tijdschrift*, 76:359-368.

- Polgár Z, Blackwell EJ, Rooney NJ. 2019. Assessing the welfare of kennelled dogs – A review of animal-based measures. *Applied Animal Behaviour Science*, 213(2019):1-13.
- Poulsen A, Lisle AT, Phillips CJC. 2010. An Evaluation of a Behaviour Assessment to Determine the Suitability of Shelter Dogs for Rehoming. *Veterinary medicine international*, 2010(2-3):523781.
- PRISMA: Transparent Reporting of Systematic Reviews and Meta-Analysis. 2020. <http://www.prisma-statement.org/> (Zugriff 01.09.2020)
- Protopopova A, Hauser H, Goldman KJ, Wynne CDL. 2018. The effects of exercise and calm interactions on in-kennel behavior of shelter dogs. *Behavioural Processes*, 146(2018):54-60.
- Protopopova A, Mehrkam LR, Boggess MM, Wynne CDL. 2014. In-Kennel Behavior Predicts Length of Stay in Shelter Dogs . <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281133/pdf/pone.0114319.pdf> (Zugriff 28.12.2020).
- Protopopova A. 2016. Effects of sheltering on physiology, immune function, behavior, and the welfare of dogs. *Physiology & Behavior*, 159(2016):95-103.
- Pschyrembel Online, 2020 Walter de Gruyter GmbH. <https://www.pschyrembel.de/> (Zugriff 31.12.2020)
- Rayment DJ, De Groef B, Peters RA, Marston LC. 2015. Applied personality assessment in domestic dogs: Limitations and caveats. *Applied Animal Behaviour Science*, 163(2):1-18.
- Righi C, Menchetti L, Orlandi R, Moscati L, Mancini S, Diverio S. 2019. Welfare Assessment in Shelter Dogs by Using Physiological and Immunological Parameters. *Animals (Basel)*, 9(6):340.
- Rollin B. 2015. An ethicist's commentary on shelter reluctant to adopt. *The Canadian veterinary journal*, 56(7):652-653.

- Sampaio RAG, de Figueiredo Martins YN, Barbosa FMS, Franco CIQ, Kobayashi MD, Talieri IC. 2019. Behavioral assessment of shelter dogs submitted to different methods of environmental enrichment. *Ciência Rural*, 49(1).
- Santesso N, Maxwell L, Tugwell PS, Buchbinder R, Johnston R. 2006. Systematic reviews from the Cochrane Musculoskeletal Group. *The Journal of the Canadian Chiropractic Association*, 50:238-43.
- Schmorrow A. 2020. Behavior assessment and modification: what is possible, what isn't? [Vortrag]. [https://www.hshv.org/wp-content/uploads/2020/01/VIS\\_Behavior-Assessment-and-Mod\\_Jan-2020\\_Final.pdf](https://www.hshv.org/wp-content/uploads/2020/01/VIS_Behavior-Assessment-and-Mod_Jan-2020_Final.pdf) (Zugriff 08.09.2020).
- Schoenfeld-Tacher R, Kogan LR, Carney PC. 2019. Perception of Cats: Assessing the Differences Between Videos and Still Pictures on Adoptability and Associated Characteristics. *Frontiers in Veterinary Science*, 6(2019):87.
- Segurson SA, Serpell JA, Hart BL. 2005. Evaluation of a behavioral assessment questionnaire for use in the characterization of behavioral problems of dogs relinquished to animal shelters. *Journal of the American Veterinary Medical Association*, 227(11):1755-1761.
- Seo MG, Kwon OD, Kwak D. 2020. Molecular detection and phylogenetic analysis of canine tick-borne pathogens from Korea. *Ticks and Tick-borne Diseases*, 11(2):101357.
- Serpell JA. 2020. Canine Behavioral Assessment & Research Questionnaire. <https://vetapps.vet.upenn.edu/cbarq> (Zugriff 14.09.2020)
- Serpell JA, Hsu Y. 2001. Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science*, 72(2001):317-364.
- Shabelansky A, Dowling-Guyer S, Quist H, D'Arpino S, McCobb E. 2015. Consistency of shelter dogs' behavior toward a fake versus real stimulus dog during a behaviour evaluation. *Applied Animal Behaviour Science*, 163(2):158-166.

- Sheppard G, Mills D. 2002. The development of a psychometric scale for the evaluation of the emotional predispositions of pet dogs. *International Journal of Comparative Psychology*, 15:201-222.
- Shore ER, Burdsal C, Douglas DK. 2008. Pet owners' views of pet behavior problems and willingness to consult experts for assistance. *Journal of applied animal welfare science*, 11(1):63-73.
- Sigma Plus Statistiek. 2021. <https://www.spss-tutorials.com> (Zugriff 16.04.2021)
- Simmons KE, Hoffman CL. 2016. Dogs on the Move: Factors Impacting Animal Shelter and Rescue Organizations' Decisions to Accept Dogs from Distant Locations. *Animals (Basel)*, 6(2):11.
- Stephen JM, Ledger RA. 2007. Relinquishing dog owners' ability to predict behavioural problems in shelter dogs post adoption. *Applied Animal Behaviour Science*, 107:88-99.
- Stephen JM, Ledger RA. 2005. An Audit of behavioral indicators of poor welfare in kennelled dogs in the United Kingdom. *Journal of applied animal welfare science*, 8(2):79-96.
- Sternberg S. 2017. *Assessing Aggression Thresholds in Dogs: Using the Assess-A-Pet Protocol to Better Understand Aggression*. Wenatchee, Washington USA : Dogwise Publishing.
- Sticco M, Trentini R, Lucidi P. 2011. Qualified "in shelter" dogs' evaluation and training to promote successful dog-human relationships. *Open Journal of Animal Sciences*, 1(3):135-144.
- Svartberg K. 2002. Shyness-boldness predicts performance in working dogs. *Applied Animal Behaviour Science* 79(2002):157-174.
- Svartberg K, Forkman B. 2002. Personality traits in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science* 79(2002):133-155.



- Svartberg K. 2005. A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science*, 91(2005):103-128.
- Svartberg K. 2021. The hierarchical structure of dog personality in a new behavioural assessment: A validation approach. *Applied Animal Behaviour Science*, 238(2021):105302.
- Syzmanski T, Casey RJ, Johnson A, Cano A, Albright D, Seivert NP. 2018. Dog Training Intervention Shows Social-Cognitive Change in the Journals of Incarcerated Youth. *Frontiers in veterinary science*, 5(2018):302.
- Tamimi N, Jamshidi S, Serpell JA, Mousavi S, Ghasempourabadi Z. 2015. Assessment of the C-BARQ for evaluating dog behavior in Iran. *Journal of Veterinary Behavior*, 10(1-2):36-40.
- Taylor KD, Mills DS. 2007. The effect of the kennel environment on canine welfare: A critical review of experimental studies. *Animal welfare*, 16(4):435-447.
- Taylor KD, Mills DS. 2006. The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior*, 1(2006):94-108.
- Thielke LE, Udell MAR. 2019. Evaluating Cognitive and Behavioral Outcomes in Conjunction with the Secure Base Effect for Dogs in Shelter and Foster Environments. *Animals (Basel)*, 9(11):932.
- Thiesen-Moussa D, Hettwer A, Hackbarth H. 2018. Der Niedersächsische Wesenstest: Ergebnisse des Testens der Gefährlichkeit von Hunden. *Berliner und Münchener Tierärztliche Wochenschrift*, 2018(131).
- Titulaer M, Blackwell EJ, Mendl M, Casey RA. 2013. Cross sectional study comparing behavioural, cognitive and physiological indicators of welfare between short and long term kennelled domestic dogs. *Applied Animal Behaviour Science*, 147(1-2):149-158.
- Turner P, Berry J, Macdonald S. 2012. Animal shelters and animal welfare: Raising the bar. *The Canadian veterinary journal*, 53(8):893-896.

- Valsecchi P, Barnard S, Stefanini C, Normando S. 2011. Temperament test for re-homed dogs validated through direct behavioural observation in shelter and home environment. *Journal of Veterinary Behavior*, 6(3):161-177
- Valsecchi P, Barnard S, Stefanini C, Normando S. 2009. Validation of a new Temperament test as a practical tool for adoptions of sheltered dogs. *Journal of Veterinary Behavior*, 4(2):75-76.
- van der Borg JAM, Netto WJ, Planta DJU. 1991. Behavioural testing of dogs in animal shelters to predict problem behaviour. *Applied Animal Behaviour Science*, 32(2-3):237-251.
- van der Borg JAM, Beerda B, Ooms M, de Souza AS, van Hagen M, Kemp B. 2010. Evaluation of behaviour testing for human directed aggression in dogs. *Applied Animal Behaviour Science*, 128(2010):78-90.
- Viggiano E, Marshall-Pescini S, Petak I, Prato-Previde E, Natoli E, Valsecchi P. 2009. Are cognitive skills good indicators of welfare in shelter dogs?. *Journal of Veterinary Behavior Clinical Applications and Research*, 4(2):79.
- Walker SL. 2014. Human and canine personality assessment instruments to predict successful adoptions with shelter dogs [Dissertation]. West Lafayette: Purdue University
- Walker JK, Dale AR, D'Eath RB, Wemelsfelder F. 2016. Qualitative Behaviour Assessment of dogs in the shelter and home environment and relationship with quantitative behaviour assessment and physiological responses. *Applied Animal Behaviour Science*, 184(2016):97-108.
- Weimer SL, Johnson AK, Sunday P, Colvin T. 2010. Independent Study 490A: Canine Behavior Evaluations: Can it Predict Canine Aggression and Therefore Improve Subsequent Adoption Rates?. *Iowa State University Animal Industry Report*, 7(1).
- Weiss E, Mohan-Gibbons H. 2012. Behavior Evaluation, Adoption, and Follow-Up. In: Miller L, Zawistowski S, Hrsg. *Shelter Medicine for Veterinarians and Staff*. Zweite Aufl. John Wiley & Sons, Inc, 529-539.

- Weiss E. 2002. Selecting shelter dogs for service dog training. *Journal of applied animal welfare science*, 5(1):43-62.
- Weiss E, Greenberg G. 1997. Service dog selection tests: Effectiveness for dogs from animal shelters. *Applied Animal Behaviour Science*, 53(1997):297-308.
- Wells DL, Hepper PG. 2000. Prevalence of behaviour problems reported by owners of dogs purchased from an animal rescue shelter. *Applied Animal Behaviour Science*, 69(2000):55-65.
- Wiener P, Haskell MJ. 2016. Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior*, 16(2016):81-85.
- Wojciechowska JI, Hewson CJ. 2005. Quality-of-life assessment in pet dogs. *Journal of the American Veterinary Medical Association*, 226(5):722-728.
- Wojciechowska JI, Hewson CJ, Stryhn H, Guy NC, Patronek GJ, Timmons V. 2005. Development of a discriminative questionnaire to assess nonphysical aspects of quality of life of dogs. *American Journal of Veterinary Research*, 66(8):1453-1460.
- Wojciechowska JI, Hewson CJ, Stryhn H, Guy NC, Patronek GJ, Timmons V. 2005. Evaluation of a questionnaire regarding nonphysical aspects of quality of life in sick and healthy dogs. *American Journal of Veterinary Research*, 66(8):1461-1467.
- Wright HF, Mills DS, Pollux PMJ. 2011. Development and Validation of Psychometric Tool for Assessing Impulsivity in the Domestic Dog (*Canis familiaris*). *International Journal of Comparative Psychology*, 24:210-225.
- Yeates JW, Main DCJ. 2008. Assessment of positive welfare: A review. *The Veterinary Journal*, 175(3):293-300.
- Zilocchi M, Ogi A, Bedini M, Mengoli M. 2015. Assessing adoptability in shelter dogs through a behavioral test. *Dog Behavior*, 2015(1):16-22.

## 9 Tabellen-/ Abbildungsverzeichnis

Tab. 1	Tabelle nach PATRONEK et BRADLEY, 2016
Tab. 2	Kriterien und Suchbegriffe der Literaturrecherche zu Wesenstests von Hunden im Tierheim
Tab. 3	Übersicht der inkludierten Studien
Tab. 4	Einteilung der Studien in Evidenzklassen
Tab. 5	Anzahl und Zeitpunkt der Follow-Up Studien – Übersicht
Tab. 6	Einteilung der inkludierten Studien in Evidenzklassen
Tab. 7	Messung der diagnostischen Testkriterien im C-BARQ-Fragebogen, angewandt an Tierheimhunden
Tab. 8	Messung der diagnostischen Testkriterien im Fragebogen nach STEPHEN et LEDGER
Tab. 9	Messung der diagnostischen Testkriterien im MCPQ-R-Fragebogen, angewandt an Tierheimhunden
Tab. 10	Übersicht über die untersuchten diagnostischen Kriterien in den Studien mit Tierheimbezug
Tab. 11	Messparameter, welche in den Fragebögen berücksichtigt wurden
Tab. 12	Statistische Validitätskriterien (die Beurteilung der prognostischen Validität ist nur dann möglich, wenn eine Follow-Up-Studie stattfand)
Tab. 13	Messung der diagnostischen Testkriterien im Wesenstest nach VAN DER BORG et al.
Tab. 14	Messung der diagnostischen Testkriterien im Ethotest© und Ethotest-R
Tab. 15	Messung der diagnostischen Testkriterien im Wesenstest nach FRANK
Tab. 16	Messung der diagnostischen Testkriterien im Wesenstest von Dog's Trust
Tab. 17	Messung der diagnostischen Testkriterien im Wesenstest nach KLAUSZ et al.

Tab. 18	Messung der diagnostischen Testkriterien im Temperament Test
Tab. 19	Messung der diagnostischen Testkriterien im Wesenstest nach ZILOCCHI et al.
Tab. 20	Messung der diagnostischen Testkriterien im Wesenstest nach Assess-A-Pet
Tab. 21	Messung der diagnostischen Testkriterien im Wesenstest nach SAFER™
Tab. 22	Messung der diagnostischen Testkriterien im Wesenstest der ASPCA®
Tab. 23	Messung der diagnostischen Testkriterien im Wesenstest nach Match-Up
Tab. 24	Messung der diagnostischen Testkriterien im Wesenstest der RSPCA
Tab. 25	Messung der diagnostischen Testkriterien im Wesenstest nach B.A.R.K.
Tab. 26	Übersicht über die untersuchten diagnostischen Kriterien in den Studien
Tab. 27	Messparameter, welche in den Wesenstests berücksichtigt wurden
Tab. 28	erfüllte Kriterien zu Praktikabilität und Objektivität
Tab. 29	Übersicht über Zeitpunkt des Wesenstests im Tierheim
Tab. 30	Test-Retest-Reliabilität der Wesenstests
Tab. 31	Inter-Rater-Reliabilität der Wesenstests
Tab. 32	Intra-Rater-Reliabilität der Wesenstests
Tab. 33	Übersicht über prognostische Validität von Wesenstests
Tab. 34	Übersicht über die untersuchten diagnostischen Kriterien in den Studien
Tab. 35	Übersicht über die untersuchten Messparameter
Tab. 36	Vierfeldertafel zu Aggression gegenüber Erwachsenen (VAN DER BORG et., 19
Abb. 1	Flussdiagramm zur Darstellung des Vorgehens bei der Literaturrecherche
Abb. 2	Gedankenexperiment nach PATRONEK und BRADLEY (2016)
Abb. 3	Grafik zur Sinnhaftigkeit von Wesenstests nach PATRONEK et BRADLEY, 2016